



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS EXACTAS
DEPARTAMENTO DE FISICA

Trabajo de Tesis Doctoral: *Aplicación de técnicas ópticas sobre tejidos y fluidos corporales para el diagnóstico no invasivo de enfermedades oncológicas*

Tesista: *Lic. María Agustina Corti*

Director/a: *Dr. Mario Garavaglia*

Codirector/a: *Dr. Fernando Sebastián García Einschlag*

Año: *2020*

*A Matias, Morena y Charo
que aparecieron en mi vida en el transcurso de esta Tesis
y la completaron para siempre.*

ÍNDICE

Agradecimientos	1
Estructura del trabajo	3
Capítulo 1. Introducción	5
1.1 Introducción a las biopsias ópticas	5
1.1.1 Métodos no asociados a imágenes: Espectroscopia	6
1.2 Espectroscopia de fluorescencia para el diagnóstico biomédico	8
1.2.1 Principios de espectroscopia de fluorescencia	9
1.2.2 Fluoróforos endógenos y exógenos	15
1.2.3 Espectroscopia de fluorescencia en medios turbios	21
1.2.4 Espectroscopia de fluorescencia en el análisis de tejidos biológicos	24
1.2.5 Espectroscopia de fluorescencia en el análisis de fluidos biológicos	26
Parte I Aplicación de técnicas ópticas sobre tejidos para el diagnóstico no invasivo de enfermedades	30
Capítulo 2. Materiales y métodos empleados para el estudio de tejidos	31
2.1 Espectroscopia de fluorescencia sobre tejidos	31
2.2 Diagnóstico fotodinámico	34
2.3 Terapia fotodinámica	34
2.4 Termografía IR	42
Capítulo 3. Biopsias ópticas de tejidos – casos estudiados	45
Caso 1. Espectros típicos de piel sana	45
Caso 2. Verrugas comunes o vulgares	47
Caso 3. Angioma plano (EEm+Termografía IR)	49
Caso 4. Proceso inflamatorio crónico liquenoide	52
Caso 5. Condilomas – VPH6	54
Caso 6. Lesión en paladar duro – VPH6	56
Caso 7. Epiteliomas basocelulares (DFD,TFD)	59
Caso 8. Adenocarcinoma de parótida (EEm+TFD+IR)	62
Caso 9. Adenocarcinoma de colon – (TFD por fibrocolonoscopia)	67
Parte II Evaluación de la utilización de la autofluorescencia de muestras de orina en la detección de enfermedades neoplásicas	71
Capítulo 4. Materiales y métodos empleados para el estudio de fluorescencia de orinas	73
4.1 Recolección y tratamiento de muestras previo a la obtención de información espectroscópica	73
4.1.1 Obtención de las muestras	73
4.1.2 Pre-tratamiento de las muestras	74
4.1.3 Pre-tratamiento de los datos	76

4.2 Técnicas analíticas	76
4.2.1 Espectrofotometría UV-VIS	76
4.2.2 Espectrofluorometría UV-VIS	77
4.3 Análisis estadístico y test diagnósticos	78
4.4 Técnicas numéricas – Herramientas de análisis multivariado	88
4.4.1 Análisis de componentes principales (PCA)	88
4.4.2 Resolución Multivariada de Curvas por cuadrados mínimos alternados (MCR-ALS)	92
4.4.3 Análisis Discriminante Lineal (LDA)	97
4.4.4 Análisis paralelo de factores – PARAFAC	102
Capítulo 5. Análisis de espectros de fluorescencia de muestras de orina	113
5.1 Investigación preliminar con espectros de fluorescencia de muestras de Orina	113
5.2 Ampliación del estudio inicial	114
5.3 Análisis estadístico de los datos	121
5.4 Análisis por componentes principales	129
5.4.1 Estudio de la variabilidad intra-individuo	137
5.5 Análisis de los espectros normalizados y sus derivadas	142
5.6 Análisis por Resolución Multivariada de Curvas (MCR)	151
5.7 Análisis Discriminante Lineal (LDA)	157
Capítulo 6. Matrices de fluorescencia de muestras de Orina	167
6.1 Análisis multivariado de la fluorescencia de orina	167
6.2 Modelo PARAFAC de las MEE medidas	173
6.2.1 Determinación de la cantidad de factores	177
6.2.2 Modelo PARAFAC obtenido	183
6.3 Análisis estadístico de las MEE	186
6.3.1 Estudio de la zona espectral óptima de las MEE para desarrollar un test diagnóstico	188
Capítulo 7 Resumen, discusión y conclusiones generales	191
7.1 Discusión de los resultados de la parte I	191
7.2 Discusión de los resultados de la parte II	193
7.3 Conclusiones	199
Índice de abreviaturas utilizadas	201
Bibliografía	203

AGRADECIMIENTOS

El presente trabajo de tesis es el resultado de varios años de trabajo durante los que conté con el apoyo tanto profesional como humano de una gran cantidad de personas e instituciones a los cuales quiero agradecer:

- Al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y a la Universidad Nacional de La Plata (UNLP) por darme la posibilidad de ingresar al sistema científico, por otorgarme los medios y el espacio físico para el desarrollo de mi tesis.
- A mis directores, Dr Mario Garavaglia y Dr. Fernando S. García Einschlag. cuyos conocimientos, confianza, paciencia y calidad humana han sido fundamentales para mi formación.
- A Miguel Pasquale, por su constante y desinteresado apoyo, por su confianza y su gran ayuda en toda la segunda parte de la Tesis.
- Al Dr. Horacio Poteca por su amplia colaboración en la primer parte del presente trabajo.
- A Maru por la compañía, por las charlas de congresos, viajes, cursos, etc.
- A los PAO's: Marian, Fede, Eli, Bruno y Licha por haberme "adoptado" y hacerme sentir parte desde el primer día.
- A mis colegas amigos, o mejor dicho amigos colegas: Rusa, Martin, Vane, Dieguito y Ger.
- A Claudia Rodríguez Torres, Laura Damonte y Sergio Sciutto, quienes me escucharon y supieron aconsejarme siempre.
- A mi mamá, a mis hermanos, cuñados, a toda la familia quienes me apoyaron en todo momento y hasta colaboraron con sus muestras.
- A Mati por ayudarme a ser, y acompañarme en todo. Sobre todo el último tiempo tuvo que poner todo de sí para que yo pueda lograr esto.
- A More y a Charo que transitaron esta tesis desde que estaban en mi panza.
- A todos los que me crucé a lo largo del camino de esta Tesis que sumaron su granito de arena e hicieron que fuera posible.
- Y finalmente a Seba, a quien en realidad nunca me alcanzarán las palabras para agradecerle... La pasión por lo que hace contagia, su dedicación es admirable, y como si todo eso fuera poco su calidad humana es excelente. Pasaron más de 20 años desde que en el secundario me enseñó que no se pueden sumar chanchos con pajaritos y que hoy me dedique a esto sin suda es gracias a él.

Estructura del trabajo

El presente trabajo de Tesis consiste en investigaciones tendientes a evaluar la factibilidad del empleo de métodos ópticos no invasivos como herramientas para el diagnóstico y monitoreo del tratamiento de enfermedades oncológicas. En este sentido, la expresión “métodos ópticos” se refiere a aquellas técnicas que poseen la capacidad de medir las interacciones entre la energía radiante y la materia. En el presente estudio se trabajó principalmente con la aplicación de la espectroscopia de fluorescencia UV/VIS como herramienta de diagnóstico en aplicaciones clínicas.

El trabajo comienza con un Capítulo introductorio, donde se plantea la importancia de desarrollar un método de diagnóstico confiable y no invasivo de enfermedades oncológicas, se comenta el estado de situación actual, se definen las biopsias ópticas, y se presenta un resumen de los principios de la espectroscopia óptica de fluorescencia. Además, en este mismo Capítulo se presentan las propiedades de fluorescencia endógena de moléculas presentes en células, tejidos y fluidos corporales, y las propiedades de fluorescencia de algunas moléculas exógenas que se emplean como agentes de contraste. Posteriormente se plantea una discusión del efecto de la absorción y la dispersión en la espectroscopia de fluorescencia de medios turbios y los métodos para eliminar este efecto. Finalmente se realiza una presentación de la espectroscopia óptica de fluorescencia aplicada al análisis de tejidos y luego de fluidos biológicos.

El presente trabajo de Tesis se divide luego en dos partes: la primera describe los estudios acerca de las denominadas “biopsias virtuales” realizadas sobre tejidos, y la segunda presenta investigaciones relacionadas con la evaluación de la utilización de la autofluorescencia (fluorescencia debida a fluoróforos endógenos) de muestras de orina para la detección de enfermedades neoplásicas.

El objetivo de la Parte I, que comprende los Capítulos 2 y 3, fue evaluar el potencial de observaciones ópticas “in vivo” para el diagnóstico de neoplasias de tejido humano empleando técnicas de fluorescencia sobre la zona tumoral y sus adyacencias. También se emplearon imágenes de termografía IR con fines diagnósticos para medir el incremento de temperatura asociado a neovascularizaciones en tejidos patológicos. Así, en el Capítulo 2 se analizan los requisitos de instrumentación para la espectroscopia de fluorescencia de tejidos, el diagnóstico fotodinámico y la terapia fotodinámica. Asimismo se presentan los aspectos básicos de la termografía IR. En el Capítulo 3 se presentan nueve casos de biopsias ópticas realizadas mediante espectroscopia óptica de fluorescencia, (tanto por autofluorescencia como por diagnóstico fotodinámico), comparando la fluorescencia natural de tejidos sanos y de tejidos patológicos. Además, se muestran los casos en los que se utilizaron imágenes termográficas (termografía IR) para contribuir al diagnóstico médico.

En la Parte II del trabajo, desarrollada en los Capítulos 4 a 6, se presentan las investigaciones realizadas con el fin de evaluar la factibilidad de diseñar un método sencillo, rápido e incruento para la detección de actividad oncogénica mediante el análisis de la autofluorescencia de muestras de orina. El objetivo de esta parte del trabajo fue contribuir al desarrollo de un procedimiento de diagnóstico que, a partir del análisis de la autofluorescencia de la orina, pueda ser utilizado como método de screening para la detección de enfermedades neoplásicas o bien para evaluar el progreso de las mismas.

En el Capítulo 4 se presentan los materiales y métodos utilizados en la segunda parte de la Tesis. En él se detalla el procedimiento seguido para la recolección, preparación y conservación de las muestras, de manera de optimizar las condiciones experimentales para la caracterización de la orina mediante el empleo de espectroscopias de absorción y de fluorescencia molecular en la región UV-visible del espectro electromagnético. Asimismo, se describen tanto el equipamiento y los métodos analíticos empleados para caracterizar el comportamiento espectroscópico de las muestras de orina como las técnicas numéricas utilizadas para el análisis estadístico de los resultados. Por último, en ese mismo Capítulo, se estudian aspectos esenciales relacionados con la evaluación de los test diagnósticos.

El Capítulo 5 corresponde al análisis de los espectros de fluorescencia de la muestras de orina, mientras que en el Capítulo 6 se presenta el análisis y la interpretación de las matrices de excitación y emisión de fluorescencia. En ambos Capítulos se evalúa el potencial de diferentes técnicas de análisis multivariado para el procesamiento de la información relevante para fines diagnósticos y de métodos estadísticos tanto para la interpretación cuantitativa de los datos espectroscópicos como para el desarrollo de test que permitan detectar situaciones patológicas o monitorear su evolución.

Por último el Capítulo 7 presenta las discusiones de ambas partes y las conclusiones generales del presente trabajo de Tesis.

Capítulo 1 Introducción

1.1 Introducción a las biopsias ópticas.

El término “biopsia” proviene del griego bíos (βίος) 'vida' y opsía (οψία) 'vista'. Según la Real Academia española, la biopsia se define como “Extracción y examen de una muestra de tejido tomada de un ser vivo, con fines diagnósticos”. En concordancia con ésta, la definición generalmente empleada en el ámbito médico se refiere a ella como “muestra de tejido extirpada de un ser vivo, con fines diagnósticos”. [1] De estas definiciones surge que la realización de una biopsia incluye dos acciones: 1) la toma o extracción de una muestra de tejido o de células del organismo y 2) su posterior análisis y estudio al microscopio. Cualquiera que sea la técnica empleada para la obtención de una muestra de tejido, que es la esencia de la biopsia, queda claro que en todos los casos el procedimiento es al menos invasivo, y en algunos casos cruento. Es decir, supone romper la integridad del cuerpo y causarle una herida de mayor o menor entidad según la técnica utilizada. Es por ello que convencionalmente asociamos el término “biopsia” a un procedimiento médico invasivo por el cual se toma una muestra de tejido humano y se remite al laboratorio de patología para su análisis con fines diagnósticos.

Las biopsias ópticas (BO) o biopsias “virtuales” se definen como aquellas técnicas que utilizan la energía de la luz para obtener información de la estructura de las células tumorales y de sus funciones en los tejidos sin ser disruptiva para los mismos. [2] Este tipo de metodología es una forma no invasiva de diagnóstico que realiza un análisis del tejido (en superficie, en cavidades, o en profundidad), usando técnicas como las siguientes: láser, tomografía de coherencia óptica (en inglés OCT), espectroscopias infrarroja y de fluorescencia, etc. Estas técnicas son capaces de diagnosticar sin involucrar una biopsia intrusiva. Al tejido a analizar se accede a través de la superficie del cuerpo, incluido el análisis de la propia piel, o por vía endoscópica. Está claro que, respecto de las biopsias convencionales, estas técnicas representan una ventaja en la atención de los pacientes ya que les evita tener que someterse a procedimientos invasivos.

Cabe destacar que a diferencia de en la Anatomía Patológica, donde el «estándar de oro»¹ es la histología del tejido fijado (o biopsia diferida), en las técnicas de BO no se dispone aún de un «estándar de oro». Por otro lado, una gran ventaja que poseen las imágenes generadas por BO es que éstas se obtienen en tiempo real y van acompañadas de una considerable información complementaria que permite evaluar la enfermedad *in vivo*, mientras que en el caso convencional, la obtención de resultados en general implica un procesamiento de hasta 48 horas.

¹ *Estándar de oro o test de referencia* es la técnica diagnóstica que define la presencia de la condición con la máxima certeza conocida. Se refiere a una prueba diagnóstica (de laboratorio, imagenología, patología, etc.) que es la más adecuada para hacer el diagnóstico de una enfermedad en particular.

Si bien las BO poseen todas las ventajas detalladas anteriormente respecto de las convencionales, en la mayoría de los casos, aún se requiere un exhaustivo trabajo de investigación para correlacionar los resultados de la biopsia óptica con los de la biopsia convencional antes de implementar un procedimiento diagnóstico basado en la información obtenida de las biopsias ópticas que prescindan de las biopsias convencionales.

Desde el punto de vista técnico los métodos de biopsia óptica se dividen en dos grandes grupos [2]:

- 1) **Métodos basados en imágenes** entre los que se encuentran todas las imágenes de luz coherente del tipo de la Tomografía de coherencia óptica (OCT), Imágenes de coherencia óptica (OCI), Imágenes de holografía digital (DHI), etc. Las de iluminación estructurada como la endomicroscopía confocal o las mixtas como la microscopía foto-acústica (PAM).
- 2) **Métodos no asociados a imágenes** que incluyen la espectroscopia de los tejidos (fluorescencia, reflectancia, dispersión fotónica elástica, etc.) con luz coherente o no coherente. A este conjunto de técnicas diseñadas para detectar patologías que no están basadas en imágenes sino en perfiles espectrales, también se las denomina “patología espectral”.

Cabe destacar que en un principio la expresión “biopsia óptica” se usó para mencionar a los métodos del segundo grupo, que por estudiar los tejidos en función del análisis espectral, se encuentran lejos del área de acción de los anatomopatólogos, quienes se basan en la observación de las modificaciones de la histología normal o en la morfología de las células y el tejido.

1.1.1 Métodos no asociados a imágenes: Espectroscopia

La espectroscopia es el estudio de la interacción de la radiación electromagnética con la materia. Un espectro es la representación de la distribución de las intensidades de radiación absorbida, emitida o dispersada, en función de cierta magnitud característica, como la longitud de onda, la energía o la temperatura, por ejemplo. Más allá de la magnitud que se desee analizar, en cualquier medición espectroscópica hay tres aspectos fundamentales que deben ser considerados: en primer lugar la irradiación de la muestra con radiación electromagnética; segundo la medición de la absorción, emisión espontánea (fluorescencia, fosforescencia) y/o dispersión (dispersión elástica de Rayleigh, dispersión inelástica Raman, etc.) de la radiación que interactuó con la muestra; y por último el análisis e interpretación de estas medidas.

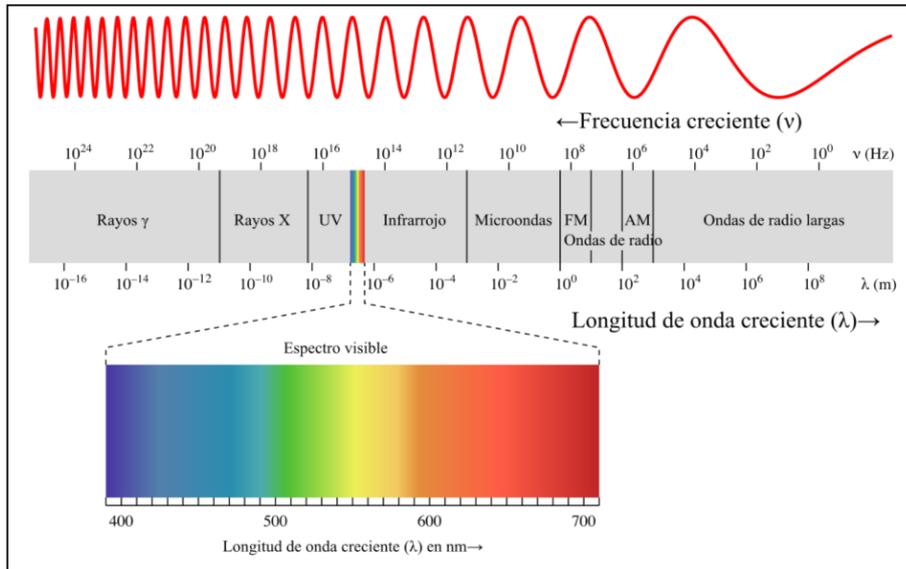


Figura 1.1 Espectro electromagnético.

En particular, la **espectroscopia óptica** estudia las interacciones de la radiación electromagnética con la materia que se produce en las longitudes de onda ultravioleta (UV), visible (VIS), infrarrojo cercano (NIR) e infrarrojo medio (MIR).[3] Tal como se detallará en la sección 1.2.3, en las regiones espectrales UV/VIS (longitud de onda entre 200 y 700 nm), la luz sólo puede penetrar desde unos pocos micrones a centímetros de profundidad en el medio biológico. Esto se debe a que dicho medio es altamente absorbente y es por eso que se utiliza sólo para analizar el tejido superficial. Sin embargo, en la región espectral NIR (longitudes de onda entre 700 y 900 nm) el tejido generalmente es menos absorbente y, además, el número de eventos de dispersión elástica de la luz en el tejido es aproximadamente dos órdenes de magnitud mayor que el número de eventos de absorción. Esto permite que la luz se propague a través de volúmenes de tejido que tienen una profundidad de hasta varios centímetros.

En la región espectral UV/VIS, las espectroscopias de absorción y de fluorescencia se han explorado ampliamente como herramientas de diagnóstico para el precáncer y la detección de cáncer en los epitelios superficiales de diversos órganos (colon, cérvix, bronquios, pulmones, vejiga, cerebro, esófago, cabeza y cuello, piel, conductos biliares, mamas y tejidos del estómago), y para la caracterización de la aterosclerosis en la superficie de las paredes de las arterias [4-11]. La espectroscopia de fluorescencia también se ha usado para detectar enfermedades del cerebro, como el Alzheimer *in vitro* [12]. La espectroscopia de dispersión elástica en esta región espectral también se ha evaluado para la detección de precánceres y cánceres en el epitelio de varios órganos, pero en un grado mucho menor. [13]

En la región espectral NIR, se ha utilizado la espectroscopia de absorción para detectar hemorragias cerebrales [14], oxigenación cerebral [15], actividad funcional del cerebro [16, 17] y tumores de mama [18]. Hasta la fecha, la mayoría de las aplicaciones clínicas se han concentrado en las espectroscopias de absorción, fluorescencia y dispersión elástica, porque estas mediciones se pueden obtener con una buena relación señal/ruido en tiempos de integración razonablemente cortos. Por todo lo expuesto anteriormente, es que el presente trabajo de Tesis se centra en el estudio de la espectroscopia óptica de fluorescencia UV/VIS como herramienta de diagnóstico para la detección de precánceres y cánceres in vivo.

1.2 Espectroscopia de fluorescencia para el diagnóstico biomédico.

Las técnicas de diagnóstico basadas en espectroscopia óptica tienen el potencial de vincular ciertas propiedades bioquímicas y morfológicas de los tejidos con la clínica de los pacientes. Se pueden utilizar para indagar acerca de ciertas características claves del tejido, como la tasa metabólica celular, la vascularización, la oxigenación intravascular y las alteraciones en la morfología del tejido. Luego, estas características del tejido se pueden interpretar para intentar diagnosticar algunas patologías como el crecimiento precanceroso y canceroso. Además, estas técnicas poseen la gran ventaja de ser rápidas, no invasivas y cuantitativas. Es por todo esto que la espectroscopia óptica tiene el potencial de representar un paso importante hacia los avances en aplicaciones tanto diagnósticas como terapéuticas.

Para poder analizar e interpretar la información espectroscópica, es fundamental conocer sus fundamentos teóricos, por ello la siguiente sección se refiere específicamente al fenómeno de fluorescencia. Mientras tanto, aquí se presenta una breve descripción desde el punto de vista de su aplicación clínica.

La espectroscopia óptica tiene la capacidad de sondear los niveles de energía de las moléculas. Estos niveles representan estados energéticos característicos de cada molécula, y están estrechamente relacionados con su estructura molecular. Al estado de energía más bajo de una molécula se lo llama estado fundamental, mientras que los estados de mayor energía se denominan estados excitados. La energía de las moléculas posee varias contribuciones originadas en distintos tipos de movimiento (electrónico, vibracional, rotacional, traslacional y aquellas asociadas con el espín nuclear y electrónico). En el régimen óptico, que es el estudiado en el presente trabajo de Tesis, los niveles de energía de interés son aquellos que están asociados con transiciones vibracionales y electrónicas.

La separación entre los niveles de energía vibracional está determinada por la masa de los átomos y la "flexibilidad" de los enlaces químicos que los unen. En cambio, la separación entre los niveles de energía electrónica, que es mayor que la vibracional, ocurre cuando los electrones se desplazan de una región de una molécula a otra.

Debido a la agitación térmica, a cualquier temperatura finita, las moléculas se distribuyen entre los niveles de energía disponibles. [3] Dicha distribución depende de la temperatura absoluta (T) y de la separación entre los niveles de energía. A una temperatura dada, el número de moléculas en un nivel superior (n_{\uparrow}) en relación con un nivel inferior (n_{\downarrow}) está dado por la distribución de Boltzmann, como se muestra en la Ecuación (1), donde k es la constante de Boltzmann ($1,380649 \times 10^{-23} \text{ JK}^{-1}$):

$$\frac{n_{\uparrow}}{n_{\downarrow}} = e^{-\Delta E/kT}. \quad (1.1)$$

Cuando radiación electromagnética incide sobre una molécula, se produce un aumento de la probabilidad de transición tanto desde un nivel de energía más bajo a un nivel de energía más alto, como así también de que ocurra en el sentido contrario. En consecuencia, la absorción neta o la transición a un nivel de energía más alto pueden ocurrir sólo si la diferencia entre las poblaciones de los niveles de energía en cuestión es importante, siendo generalmente significativamente mayor la población del estado de menor energía. El cálculo de la población de niveles de energía a temperatura ambiente ha demostrado que para las separaciones asociadas a los niveles de energía vibracionales la relación $n_{\uparrow}/n_{\downarrow}$ es del orden de 10^{-3} y para las separaciones correspondientes a los niveles de energía electrónicos de 10^{-21} .

1.2.1 Principios de espectroscopia de fluorescencia.

Cuando incide radiación electromagnética sobre las moléculas, si el contenido energético de la radiación incidente se corresponde con la diferencia de energía entre un nivel energético ocupado por un electrón y otro vacío, puede producirse una transición electrónica entre el estado fundamental y el correspondiente estado excitado (figura 1.2). Dado que, en general, los estados excitados son inestables, las moléculas tienden a retornar rápidamente al estado de mínima energía. El sistema puede volver al estado fundamental liberando el exceso de energía mediante diversos mecanismos de relajación. Dos de los más importantes de estos mecanismos son la relajación no radiante (relajación vibracional y conversión interna) y la relajación fluorescente, que es un mecanismo de relajación radiativo, al igual que la fosforescencia. [19]

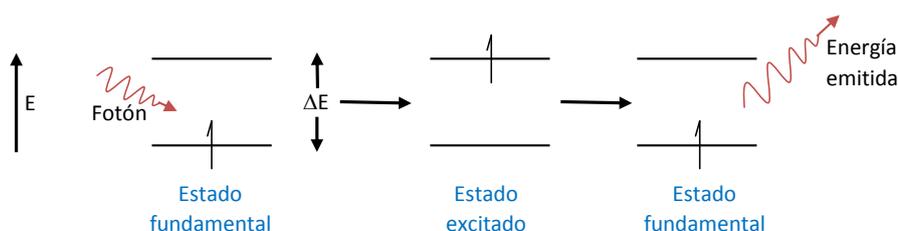


Figura 1.2. Transiciones electrónicas inducidas por la absorción y emisión de energía radiante.

La relajación vibracional, tiene lugar durante las colisiones entre moléculas excitadas y las moléculas de su entorno, y la ganancia de energía vibracional de éste último se refleja en un ligero incremento de la temperatura. Este es un proceso tan eficiente que el tiempo de vida promedio de un estado vibracional excitado es de 10^{-15} s, aproximadamente. Otro tipo de relajación, llamado conversión interna [19], es debido a la relajación no radiativa entre el nivel vibracional inferior de un estado electrónico excitado y el nivel vibracional superior de otro estado electrónico de menor energía.

A diferencia de estos dos mecanismos, la fluorescencia es un fenómeno físico en el cual las moléculas relajan al estado basal, liberando su exceso de energía emitiendo fotones. Una de las características más atractivas de los métodos de fluorescencia es su sensibilidad inherente, la cual es, con frecuencia, de uno a tres órdenes de magnitud mayor que la de la espectroscopia de absorción. No obstante, los métodos de fluorescencia se aplican mucho menos que los métodos de absorción debido al número relativamente limitado de sistemas químicos que presentan tasas de emisión medibles. [20]

En la figura 1.3 se observan las transiciones electrónicas que pueden ocurrir durante la relajación de moléculas excitadas. En el esquema S_0 representa al estado electrónico fundamental, mientras que S_1 , S_2 , S_3 y T_1 representan estados excitados. IC_{32} e IC_{21} indican la conversión interna de S_3 a S_2 y de S_2 a S_1 , respectivamente, F_2 , F_1 , P_1 indican los decaimientos radiativos hasta S_0 desde S_2 , S_1 y T_1 , respectivamente, y *Vib* es la relajación vibracional en los estados excitados S_2 , S_1 y T_1

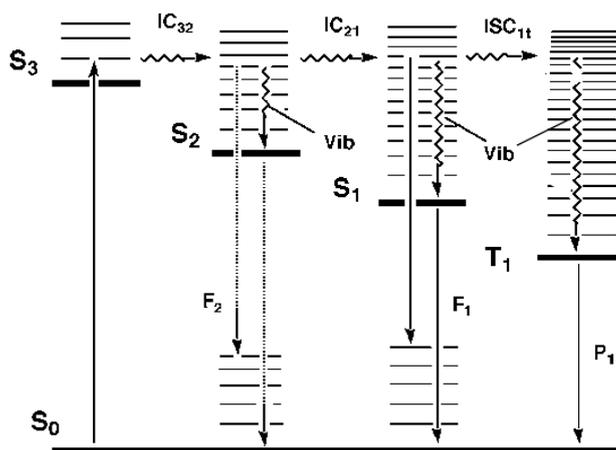


Figura 1.3 Esquema de los procesos radiativos y no radiativos de moléculas orgánicas (S:singulete, T: triplete, IC: conversión interna).

El camino más probable hacia el estado fundamental es aquel que minimiza el tiempo de vida del estado excitado. Por tanto, si la desactivación por fluorescencia es rápida con respecto a los procesos no radiativos, se observa tal emisión. Por otro lado, si

un camino de desactivación no radiativa tiene una constante de velocidad favorable, la fluorescencia no tiene lugar o es menos intensa. [21]

En las moléculas orgánicas, los electrones que participan en las transiciones electrónicas en presencia de radiación UV-Vis son los que forman parte de los enlaces (simples, dobles, triples), así como los pares de electrones no compartidos presentes en los heteroátomos (N, P, O, S, halógenos), pero no los electrones más internos, ya que estos requieren una energía mucho mayor para ser excitados. Así pues, las transiciones electrónicas que son relevantes para la absorción de la radiación en la región UV-Vis del espectro se producen desde orbitales moleculares enlazantes (σ o π) o de no enlace (n) hasta orbitales moleculares antienlazantes (σ^* o π^*). El orden de energía de las transiciones mencionadas se muestra en la figura 1.4.

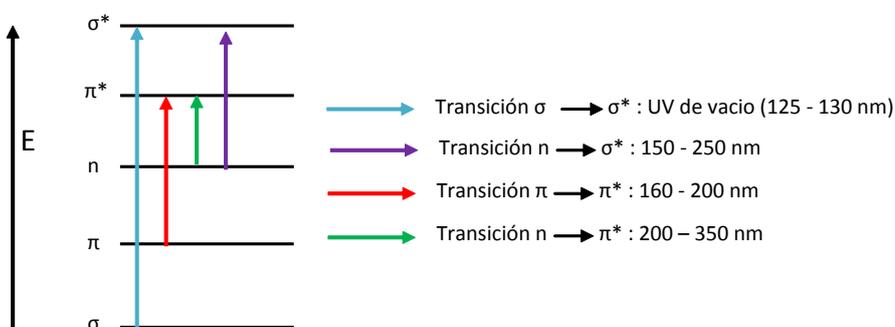


Figura 1.4. Transiciones electrónicas que son relevantes para la absorción de la radiación en la región UV-Vis del espectro.

Las bandas de fluorescencia molecular están formadas principalmente por bandas que tienen longitudes de onda más largas y, por tanto, energías menores que la banda de radiación responsable de su excitación. Este desplazamiento hacia longitudes de onda mayores, dado por la diferencia en la frecuencia (o longitud de onda) entre la luz absorbida y la luz emitida se denomina desplazamiento (o corrimiento) de Stokes [22], y puede aprovecharse estratégicamente para la obtención de imágenes basadas en la fluorescencia.

Los casos en los que la radiación absorbida es emitida sin cambio en la longitud de onda (misma energía) se conocen como radiación de resonancia o resonancia fluorescente.[22]

En general, el espectro de excitación es, muy frecuentemente, similar al espectro de absorción. Además, debido a que las diferencias de energía entre los estados vibracionales son aproximadamente las mismas, tanto para el estado fundamental como para el excitado, el espectro de excitación y el espectro de emisión o fluorescencia de un

compuesto frecuentemente, en ausencia de otras interacciones, aparecen como la imagen especular uno del otro con una sobreposición que ocurre en la línea de resonancia.[22] Los espectros de absorción y emisión de un fluoróforo² hipotético se muestran en la figura 1.5.

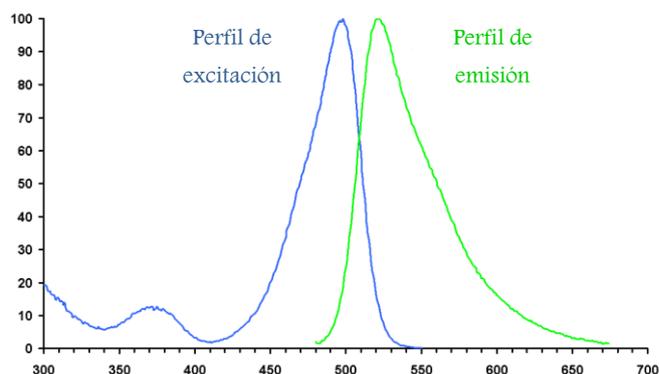


Figura 1.5 Esquema de los perfiles de excitación y de emisión de una molécula.

En líneas generales, se puede decir que la generación de fluorescencia ocurre en tres pasos:

- 1- Inmediatamente después de la excitación, se logra una suerte de “equilibrio térmico” a medida que el electrón realiza una transición no radiativa al nivel vibracional más bajo del primer estado excitado. Estrictamente hablando este equilibrio se refiere a que la población de los estados vibracionales del S_1 está de acuerdo con la temperatura del medio.
- 2- Se produce una transición *radiativa* a un nivel vibratorio excitado del estado electrónico fundamental (S_0).
- 3- El sistema experimenta una transición *no radiativa* al nivel vibracional más bajo del estado fundamental.

Por otro lado, si se produce un cruce entre sistemas, en el que en el estado excitado el espín del electrón se invierte, el tiempo para la transición radiativa del estado excitado al estado fundamental es más largo porque la transición debe ocurrir con un cambio de espín. Este estado excitado se denomina estado triplete (ver figura 1.3). En ese caso, la transición radiativa desde el estado triplete excitado se denomina fosforescencia y en general se da en una escala de tiempo mucho más larga que la fluorescencia. [23]

El interés en estudiar la absorción, la fluorescencia y la fosforescencia de tejidos radica en que dichos procesos monitorean los cambios en los niveles de energía electrónica, proporcionando valiosa información bioquímica de moléculas biológicas. En el

² Dada una molécula capaz de absorber fotones y emitir fotones de menor energía (o mayor longitud de onda), se denomina fluoróforo a la parte de esa molécula responsable de la emisión fluorescente.

presente trabajo de Tesis nos abocamos a estudiar los fenómenos de absorción y fluorescencia.

En particular el fenómeno de fluorescencia evidencia varias características generales para una molécula biológica particular. Primero, tal como se mencionó previamente, debido a las pérdidas de energía entre la absorción y la emisión que se producen como resultado de las transiciones no radiativas, la fluorescencia se produce en las longitudes de onda de emisión que siempre están desplazadas al rojo en relación con la longitud de onda de excitación. Segundo, debido a la rápida relajación de los estados superiores (S_2 , S_3 , etc.) hacia el estado S_1 mediante procesos no radiativos, las longitudes de onda de emisión son generalmente independientes de la longitud de onda de excitación.

Dos parámetros muy importantes que caracterizan la fluorescencia de una molécula son su rendimiento cuántico y su vida media. El rendimiento cuántico es una medida de la eficiencia de emisión del fluoróforo. Es simplemente la relación entre el número de fotones emitidos y el número de fotones absorbidos, y por lo tanto varía entre cero y uno. Las moléculas altamente fluorescentes tienen rendimiento cuántico cercano a la unidad mientras que las especies no fluorescentes tienen rendimiento prácticamente nulo. La vida media da una idea del tiempo promedio que la molécula biológica pasa en el estado excitado antes de regresar al estado fundamental. El rendimiento cuántico de fluorescencia y la vida media se ven afectados por una serie de factores que pueden aumentar o disminuir las pérdidas de energía. Por ejemplo, una molécula puede ser no fluorescente como resultado de una gran tasa de decaimiento no radiativo (disipación de energía al baño térmico) o una lenta tasa de decaimiento radiativo (liberación de energía en forma de radiación - fluorescencia).

La espectroscopia de fluorescencia abarca la medición y el análisis de varias características relacionadas con el rendimiento cuántico de fluorescencia y/o la vida media del estado excitado emisor de una molécula. La intensidad de fluorescencia de una molécula biológica depende de varios factores como son su concentración, su coeficiente de extinción (poder de absorción) a la longitud de onda de excitación y su rendimiento cuántico a la longitud de onda de emisión. [24]

Existen varias maneras de adquirir y de representar los datos de la emisión fluorescente. En cuanto al modo de adquisición, un espectro de emisión de fluorescencia (EEm) registra la intensidad de fluorescencia medida en un rango de longitudes de onda de emisión a una longitud de onda de excitación fija. En cambio, un espectro de excitación de fluorescencia (EEx) registra la intensidad de fluorescencia en una longitud de onda de emisión particular para un rango de longitudes de onda de excitación. Por otro lado, una matriz de excitación-emisión de fluorescencia (MEE) es un arreglo bidimensional de números, en el que se almacenan los valores de intensidad de fluorescencia en un rango

de longitudes de onda de excitación y emisión. Los espectros de emisión y excitación suelen representarse gráficamente como una línea continua en un par de ejes ortogonales, donde las ordenadas corresponden a la intensidad de fluorescencia registrada y las abscisas a la longitud de onda (o energía) de excitación/emisión correspondiente. En cambio, para representar las MEE se utilizan gráficos de contorno bidimensionales, o bien gráficos de superficie en 3 dimensiones. Una gráfica de contorno ofrece una vista bidimensional de la superficie, donde los puntos que tienen el mismo valor de intensidad de fluorescencia se conectan para producir líneas de respuestas constantes. Estas figuras se pueden representar mediante áreas sombreadas, líneas de contorno o ambas. En las gráficas de superficie 3D las longitudes de onda de excitación y emisión se representan en los ejes X e Y al igual que en las de contorno, pero estos ejes se muestran en perspectiva y sobre un tercer eje (vertical) se representa la intensidad de fluorescencia. La figura 1.6 ilustra un espectro de emisión de fluorescencia (a), un espectro de excitación de fluorescencia (b), un contorno de una MEE de fluorescencia (c) y una superficie 3D para una MEE típica (d).

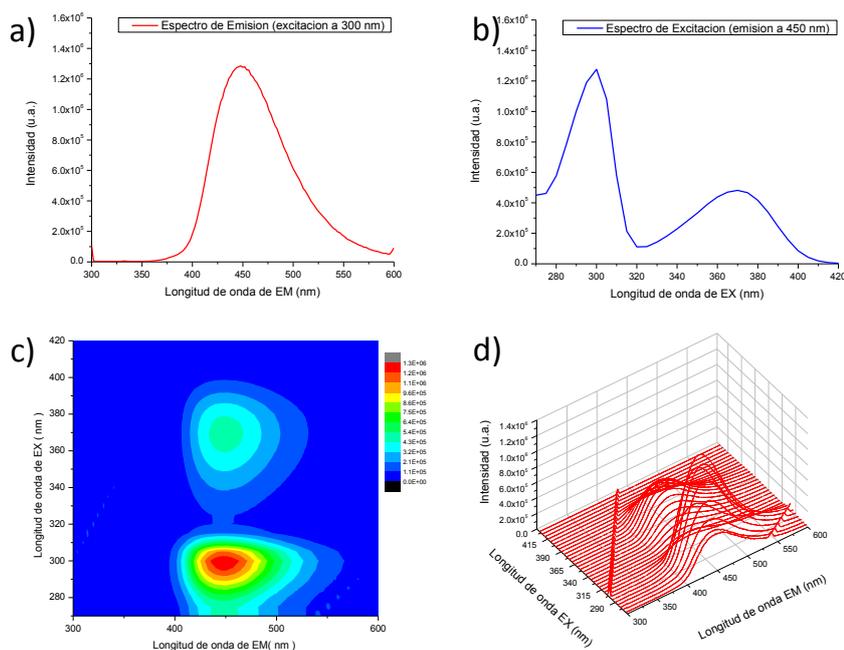


Figura 1.6 Ilustración de (a) Espectro de emisión de fluorescencia; (b) Espectro de excitación de fluorescencia; (c) Contorno de una Matriz de excitación-emisión de fluorescencia y (d) Superficie 3D de una Matriz de excitación-emisión de fluorescencia.

Cabe destacar que además de posibilitar el estudio de los espectros, y las matrices de excitación-emisión, el fenómeno de fluorescencia es también la base para modalidades de imágenes excepcionalmente potentes. Entre sus ventajas principales se destacan su alta resolución espacial y temporal, su alta sensibilidad y su cuantificabilidad. Además, la

fluorescencia es extraordinariamente sensible al entorno que rodea al fluoróforo en términos del pH, la presencia de iones y la concentraciones de oxígeno.

1.2.2 Fluoróforos endógenos y exógenos

Actualmente se cuenta con mucha información acerca de una amplia variedad de especies de interés bioquímico que presentan fluorescencia, que han sido estudiadas en profundidad. Una primera clasificación las divide de acuerdo a si la especie considerada es un fluoróforo endógeno (de origen natural) o exógeno (administrado externamente).

Fluoróforos endógenos

Dentro de las técnicas de diagnóstico basadas en el análisis de la fluorescencia, las que utilizan los fluoróforos endógenos obviamente representan la forma ideal. En esta metodología de análisis, la luz se utiliza para excitar los fluoróforos que naturalmente posee el tejido, brindando la posibilidad de obtener información sin necesidad de utilizar marcadores o fluoróforos exógenos. [25] La emisión fluorescente de este tipo de fluoróforos se denomina “fluorescencia natural” o “auto fluorescencia”.

Tal como se verá en el presente trabajo de Tesis, las diferencias en la composición química entre los distintos tipos de tejido son las que, en un futuro próximo, permitirán realizar diagnósticos en tiempo real sin extraer una muestra de tejido, ni la necesidad de incluir un fármaco que mejore el contraste. Debido a esto y al gran potencial que han demostrado los estudios iniciales, la investigación en este campo está experimentando un tremendo crecimiento.

La Tabla 1.1 presenta una lista de moléculas biológicas que exhiben fluorescencia endógena, junto con sus máximos de excitación y emisión. Estos fluoróforos endógenos incluyen aminoácidos, proteínas estructurales, enzimas y coenzimas, vitaminas, lípidos y porfirinas. Sus máximos de excitación se encuentran en el rango 250–450 nm, mientras que sus máximos de emisión se encuentran en el rango 280–700 nm, ambos comprendidos dentro de la región espectral UV/VIS.

Endogenous fluorophores	Excitation maxima (nm)	Emission maxima (nm)
Amino acids		
Tryptophan	280	350
Tyrosine	275	300
Phenylalanine	260	280
Structural proteins		
Collagen	325	400, 405
Elastin	290, 325	340, 400
Enzymes and coenzymes		
FAD, flavins	450	535
NADH	290, 351	440, 460
NADPH	336	464
Vitamins		
Vitamin A	327	510
Vitamin K	335	480
Vitamin D	390	480
Vitamin B₆ compounds		
Pyridoxine	332, 340	400
Pyridoxamine	335	400
Pyridoxal	330	385
Pyridoxic acid	315	425
Pyridoxal 5'-phosphate	330	400
Vitamin B ₁₂	275	305
Lipids		
Phospholipids	436	540, 560
Lipofuscin	340–395	540, 430–460
Ceroid	340–395	430–460, 540
Porphyrins	400–450	630, 690

FAD, flavin adenine dinucleotide; NADH, reduced nicotinamide adenine dinucleotide; AND(P)H, reduced nicotinamide adenine dinucleotide phosphate.

Tabla 1.1: Máximos de excitación y emisión de moléculas biológicas que exhiben fluorescencia endógena. Adaptada de [3].

En condiciones patológicas, las rutas metabólicas de ciertos fluoróforos endógenos se ven alteradas. En particular, en las transformaciones que ocurren con la carcinogénesis, se especula que los que más evidencian estos cambios son los aminoácidos triptófano y tirosina, las proteínas estructurales colágeno y elastina, las coenzimas NADH y FAD y las porfirinas. Dada la importancia que tienen para el presente trabajo de Tesis, a continuación se presenta una breve descripción de cada grupo de fluoróforos.

Aminoácidos. Los aminoácidos son las unidades estructurales básicas de las proteínas. Es de gran interés su estudio ya que las proteínas juegan un papel crucial en prácticamente todos los procesos biológicos. Existen tres aminoácidos con cadenas laterales aromáticas que son fluorescentes: el triptófano, la tirosina y la fenilalanina. Por debajo de 280 nm, los tres pueden ser excitados, aunque el rendimiento cuántico de emisión (ϕ_{em}) de la fenilalanina es relativamente bajo en comparación con los valores de (ϕ_{em}) asociados a triptófano y tirosina. Al excitar entre 280 y 295 nm tanto la tirosina como el triptófano son fluorescentes, sin embargo, es muy común la transferencia de energía

por resonancia de fluorescencia³ (FRET) del triptófano a la tirosina. A longitudes de onda de excitación superiores a 295 nm, sólo el triptófano es fluorescente.

Proteínas Estructurales: Colágeno Y Elastina. El colágeno comprende una familia de proteínas fibrosas, que se caracteriza por formar fibras insolubles que tienen una alta resistencia a la tracción. Es el principal componente de la matriz extracelular que está presente en cierta medida en casi todos los órganos y sirve para mantener las células juntas en unidades discretas. La elastina es el componente principal de las fibras elásticas y se encuentra en la mayoría de los tejidos conectivos junto con el colágeno y los polisacáridos. Tanto el colágeno como la elastina emiten fluorescencia al ser excitados con radiación electromagnética de la longitud de onda apropiada. La fluorescencia del colágeno tiene un máximo de excitación a 325 nm y un máximo de emisión a 400 nm, (Figura 1.7). La excitación y los máximos de emisión de la fluorescencia de la elastina ocurren a 320 y 405 nm, respectivamente, como se muestra en la figura 1.8.

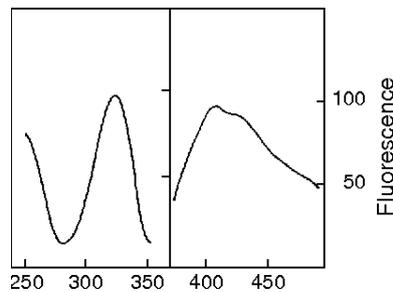
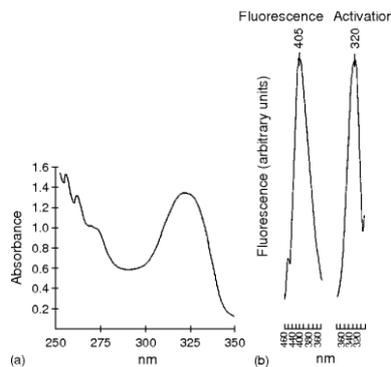


Figura 1.7 Fluorescencia del colágeno: (a) Espectro de excitación de fluorescencia con la emisión fijada en 400 nm, y (b) espectro de emisión de fluorescencia, con la excitación fijada a 325 nm. Ambos en 0,02M de buffer fosfato de potasio, pH7.4. Adaptada de [3].



³ La transferencia de energía por resonancia de fluorescencia (FRET, del inglés fluorescence resonance energy transfer), también conocida como transferencia lineal de energía, es una interacción que ocurre entre de dos moléculas fluorescentes en las que la longitud de onda de emisión de una de ellas coincide con la de excitación de la otra. Para esto es fundamental que los dos estados de excitación electrónica estén muy cercanos entre sí.

Figura 1.8 Fluorescencia de la elastina: (a) Espectro de absorción UV en 0,1M de buffer fosfato de potasio; (b) espectros de emisión y excitación de fluorescencia en 0,1M de HCl. Adaptada de [3].

Coenzimas metabólicas: nucleótidos de piridina y flavoproteínas. Los organismos vivos requieren una entrada continua de energía libre para poder realizar el trabajo mecánico en la contracción muscular y otros movimientos celulares, el transporte activo de moléculas e iones y la síntesis de macromoléculas y otras biomoléculas a partir de precursores simples. En la mayoría de los procesos metabólicos, el portador de energía libre es el trifosfato de adenosina (ATP), cuya síntesis está acoplada a la oxidación de las moléculas que sirven como combustible, tales como los carbohidratos y los ácidos grasos.

En los organismos aeróbicos, el último agente oxidante o aceptor de electrones es el oxígeno molecular. Sin embargo, los electrones no se transfieren directamente de las moléculas combustible y sus productos de descomposición al oxígeno molecular. En su lugar, estos sustratos transfieren electrones a portadores especiales llamados piridin nucleótidos y flavin nucleótidos. La forma oxidada del piridin nucleótido es la nicotinamida adenina dinucleótido (NAD), y junto con la forma oxidada del flavin nucleótido (FAD) son los principales aceptores de electrones en la oxidación de las moléculas. Al aceptar los electrones, estos portadores se reducen. Los portadores de electrones ahora reducidos (NADH) y dinucleótido de flavina y adenina (FADH₂), luego transfieren sus electrones al oxígeno molecular por medio de la cadena de transporte de electrones en la membrana interna de las mitocondrias dentro de las células. Como resultado, el oxígeno molecular los oxida a NAD y FAD. El flujo de estos electrones al oxígeno molecular es el que impulsa la síntesis de ATP.

Las formas fluorescentes de estos portadores de electrones son las formas reducidas del NAD y oxidadas del FAD, que emiten fluorescencia cuando se excitan con radiación UV y luz azul respectivamente. La Figura 1.9 muestra los espectros de excitación y emisión de fluorescencia, medidos a -196°C, del flavin nucleótido (Fp) y piridin nucleótido (PN) de mitocondrias cardiacas aisladas de palomas. La Figura 1.9 (a) muestra la excitación de Fp oxidada (FAD), y reducida (FADH₂) indicadas como "Ox" y "Red" respectivamente. La figura 1.9 (b) muestra sus espectros de emisión correspondientes. Siguiendo el mismo esquema, las Figuras 1.9 (c) y (d) muestran los espectros de excitación y emisión de PN oxidado (NAD) y reducido (NADH), indicados como "Ox" y "Red" respectivamente. Tal como puede observarse en las figuras, la intensidad de la fluorescencia de Fp es máxima cuando se oxida y mínima cuando se reduce, al contrario de lo que ocurre para PN. Entonces, la relación de la concentración de portadores de electrones oxidados y reducidos que puede ser estimada a través de la relación entre la fluorescencia de Fp y PN, proporcionando así una medida de la tasa metabólica celular.

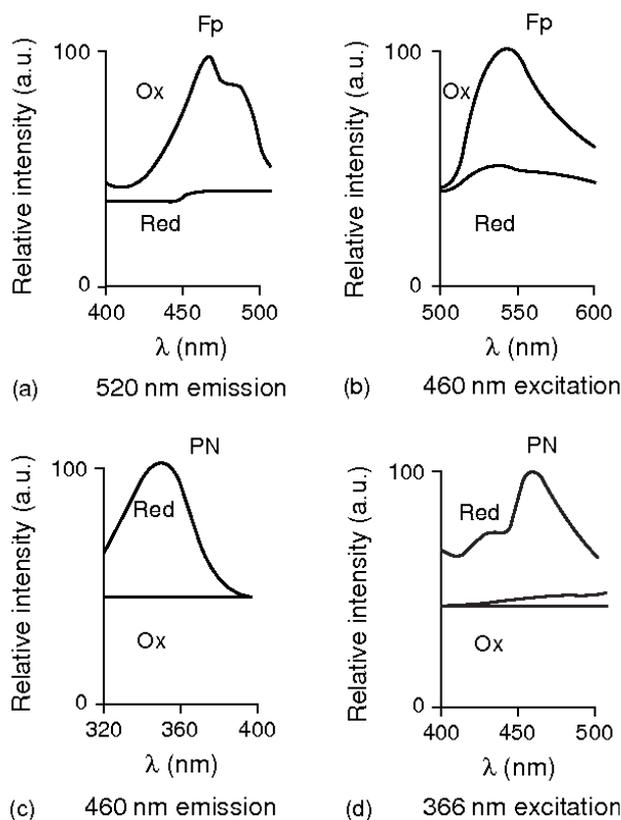


Figura 1.9 (a, b) Espectros de excitación y emisión de Fp oxidado (Ox) y reducido (Red), respectivamente. (c, d) Espectros de excitación y emisión de PN oxidado (Ox) y reducido (Red), respectivamente. Las mediciones se realizaron a partir de mitocondrias de corazón de paloma a -196°C . Adaptado de [26]

Porfirinas. La porfirina es un precursor en la biosíntesis del grupo hemo. En la figura 1.10 se muestra un esquema de la vía de biosíntesis del hemo completa. La síntesis de hemo implica ocho pasos enzimáticos que tienen lugar en las mitocondrias y el citosol. La ALA sintasa es la primera enzima mitocondrial limitante de la velocidad y forma ALA a partir de glicina y succinil Co-A. (a) y (b) representan las enzimas citosólicas ALA deshidratasa, que forma porfobilinógeno (PBG), y porfobilinógeno desaminasa, que produce dos productos: dipirrometano (un cofactor esencial) y 1-hidroximetilbilano, el tetrapirrol lineal (b). Más tarde, se sintetizan uropofirina, copropofirina y protopofirina. La protopofirina se importa a la mitocondria (c), donde la ferroquelatasa inserta un ion ferroso en la protopofirina IX (PpIX) para producir el grupo hemo (d).

Cuando son excitadas con radiación electromagnética de entre 400 y 450 nm (que corresponden a la región azul del espectro visible), las porfirinas producen fluorescencia color roja, con bandas centradas en 630 y 690-700 nm.

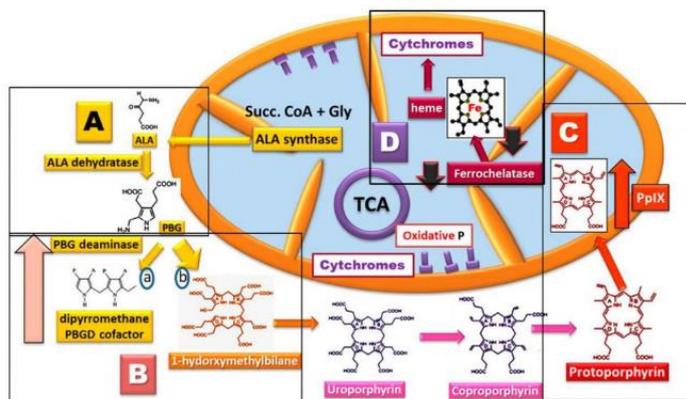


Figura 1.10 La vía de biosíntesis del grupo hemo. Adaptada de [27].

Fluoróforos exógenos

Los fluoróforos exógenos son aquellos que no se producen de forma natural dentro del tejido. Se emplean intencionalmente en determinadas aplicaciones clínicas por varias razones. La más común es proporcionar un agente de contraste que facilite ciertos diagnósticos médicos, de manera similar a los radionucleidos, que se usan como agentes de contraste en estudios circulatorios. Sin embargo, debido a la limitada penetración de la radiación electromagnética del rango óptico en los tejidos biológicos, el tipo más común de análisis de fluorescencia que se realiza *in vivo* es el de los tejidos ópticamente accesibles. Por esta razón, la mayoría de los diagnósticos de fluorescencia clínicos que utilizan fluoróforos exógenos se realizan en áreas afectadas de directa visualización como son las de la piel, el interior de cavidades como la boca, e incluso algunos sitios de acceso vía endoscópica.

La mayoría de los fluoróforos exógenos que actualmente se evalúan como agentes de contraste para la detección de precáncer y cáncer incluyen fotosensibilizadores, desarrollados originalmente para terapia fotodinámica (PDT por sus siglas en inglés)⁴. Estos fármacos generalmente muestran fuertes propiedades de fluorescencia y se ubican con preferencia en tejidos malignos. [25] Entre estos se destaca el m-tetrahidroxifenil clorina (m-THPC, Temoporfin o Foscan®), que es un derivado porfirínico sintético entre otros. Estos compuestos poseen gran absorción característica en la zona azul del espectro, y una segunda banda de absorción en la región correspondiente al rojo del espectro visible, y emiten fluorescencia entre 625 y 675 nm. Estos agentes generalmente se administran de forma sistémica y la diferencia en la acumulación de estos fluoróforos exógenos en el tejido tumoral en relación con su contraparte normal parece correlacionarse con las diferencias en la vascularización de los dos tipos de tejido.

⁴ La terapia fotodinámica se describe en detalle en el capítulo 2.

Durante los últimos años se ha introducido un concepto alternativo, que se basa en las observaciones iniciales de Ghadially y otros.[28] Ellos sugirieron, que en lugar de aplicar un fluoróforo exógeno, se use un precursor de la protoporfirina IX (PpIX) para inducir la fluorescencia en los tumores. La PpIX se caracteriza por un máximo de absorción a 405 nm y máximos de fluorescencia a 630 y 690-700 nm. Tal como se explicó previamente en esta misma sección, y se mostró en la figura 1.10, el ácido d-aminolevulínico, usualmente conocido como 5-ALA, es un precursor natural de PpIX en la ruta biosintética del grupo hemo. Normalmente, la síntesis de hemo regula la síntesis de PpIX a través de un control de retroalimentación. Sin embargo, la administración de 5-ALA pasa por alto esta retroalimentación e induce la acumulación de PpIX en el tejido tumoral.

Resumiendo, existen dos ventajas principales asociadas con el uso de 5-ALA como agente contraste: tanto el 5-ALA como el PpIX son sustancias naturalmente presentes en el cuerpo, lo que hace que el problema de toxicidad sea menos crítico; además, el fármaco puede administrarse tanto por vía oral como tópica, según sea conveniente para el caso en particular.

1.2.3 Espectroscopia de fluorescencia en medios turbios

La intensidad de fluorescencia en un medio homogéneo ópticamente diluido se puede expresar como una función lineal de la concentración de los fluoróforos presentes en ese medio. Así, la intensidad de la fluorescencia $F(\lambda_{ex}, \lambda_{em})$ a una determinada longitud de onda de excitación (λ_{ex}) y de emisión (λ_{em}), debida a los k fluoróforos presentes se define de acuerdo con la ecuación (1.2),

$$F(\lambda_{ex}, \lambda_{em}) = P_o(\lambda_{ex}) \int_0^L \sum_{k=1}^N \mu_{ak}(\lambda_{ex}) \phi_{k}(\lambda_{em}) \Omega dz \quad (1.2)$$

donde $P_o(\lambda_{ex})$ es la potencia de la luz incidente a la longitud de onda de excitación, L es la longitud de la trayectoria, Ω es la eficiencia de colección del detector y $\phi_{k}(\lambda_{em})$ es el rendimiento cuántico de emisión del fluoróforo k a la longitud de onda de emisión λ_{em} . [3] $\mu_{ak}(\lambda_{ex})$ es el coeficiente de absorción del fluoróforo k a la longitud de onda de excitación λ_{ex} , definido según la ecuación (1.3):

$$\mu_{ak}(\lambda_{ex}) = 2,303 \varepsilon(\lambda_{ex}) C \quad (1.3)$$

El coeficiente de absorción $\mu_{ak}(\lambda_{ex})$ es una función lineal del coeficiente de extinción a la longitud de onda de excitación, $\varepsilon(\lambda_{ex})$, y de la concentración C del fluoróforo k .

Aunque la espectroscopia de fluorescencia de medios homogéneos ópticamente diluidos es bien conocida, y se puede describir sencillamente con las ecuaciones 1.2 y 1.3, la interpretación de datos espectroscópicos de medios turbios, como tejidos humanos por ejemplo, es más complicada. Esto se debe principalmente a que un medio turbio, en

general es altamente absorbente y dispersante. Entonces, al analizar los datos espectroscópicos de estos medios hay que tener en cuenta varios factores entre los que se destacan la concentración de los fluoróforos presentes en el tejido, su distribución, y el entorno bioquímico/biofísico que los rodea. La importancia del entorno radica en que éste puede alterar el rendimiento cuántico, y el tiempo de vida de los fluoróforos. A modo de ejemplo, los tejidos epiteliales generalmente tienen dos subcapas primarias: un epitelio superficial y el estroma o submucosa subyacente. Los fluoróforos presentes, así como su concentración y distribución, pueden variar significativamente entre estas dos capas de tejido. Entonces, la espectroscopia de fluorescencia en este caso, depende de los fluoróforos presentes en cada subcapa de tejido. Además, como se describe a continuación, también depende de la concentración y distribución de absorbentes y dispersores no fluorescentes, dentro de cada una de las capas.

El efecto de las variables mencionadas anteriormente en la espectroscopia de fluorescencia depende también de la longitud de onda. Los fluoróforos endógenos que tienen bandas de absorción en el mismo rango de longitud de onda que la luz de excitación se excitarán y, por lo tanto, emitirán fluorescencia. Luego, las propiedades de absorción y dispersión del tejido afectarán a la luz en ambas longitudes de onda (excitación y emisión). Por lo tanto, solo aquellos fluoróforos contenidos en las capas de tejido en las que penetra la luz de excitación y de las cuales la luz emitida puede escapar de la superficie del tejido producirán una fluorescencia medible.

La absorción del tejido en las regiones UV, VIS y NIR se atribuye principalmente a la hemoglobina [29]. Aunque la absorción en el tejido es muy dependiente de la longitud de onda (ver figura 1.11) , en general tiende a disminuir con longitudes de onda crecientes. Los coeficientes de absorción del tejido varían de 0,1 a 10 000 cm^{-1} desde el NIR hasta el rango espectral UV. En consecuencia, la profundidad de penetración de la luz, que es principalmente una función de las propiedades de absorción del tejido, disminuye desde varios centímetros para el NIR hasta unos pocos cientos de micrones para el UV. Por ejemplo, en la región espectral UV, la profundidad de penetración de la luz en el tejido es de aproximadamente 225 μm a 337 nm. [30]

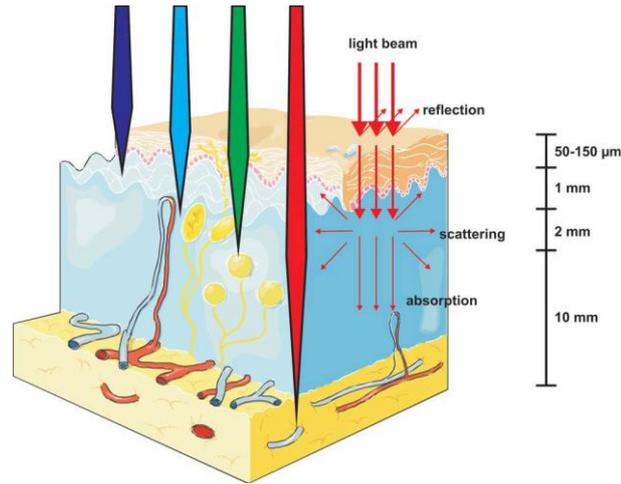


Figura 1.11 Penetración de la luz en el tejido según la longitud de onda de la luz incidente. Adaptado de [31]

La hemoglobina, que está contenida en los glóbulos rojos, sirve como portador de oxígeno en la sangre y también desempeña un papel vital en el transporte de dióxido de carbono e iones de hidrógeno. Su capacidad para unirse al oxígeno depende de la presencia de un grupo hemo. Dicho grupo consiste en una parte orgánica, un anillo de protoporfirina y un átomo de hierro. [29] La Figura 1.12 muestra los espectros de absorción de hemoglobina oxigenada y desoxigenada en la región espectral NIR.

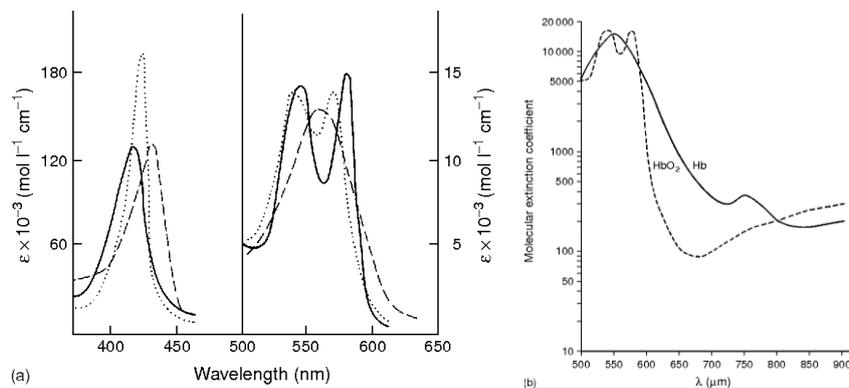


Figura 1.12 Espectros de absorción de hemoglobina oxigenada y desoxigenada sobre (a) región espectral UV / VIS (HbO_2 : línea continua; Hb : línea discontinua; banda de Soret: línea punteada) y (b) región espectral NIR. Adaptado de [3].

Tal como se mencionó previamente, los espectros se caracterizan por la continua disminución de la absorción de la hemoglobina desoxigenada y por el crecimiento de la absorción de la hemoglobina oxigenada entre 700 y 900 nm.

Otra variable que puede afectar tanto la intensidad como la forma de los espectros en las mediciones de fluorescencia del tejido es la geometría de iluminación y recolección

de la luz emitida. Esto puede atribuirse al hecho de que, aunque la fluorescencia se genera isotrópicamente a partir de los fluoróforos dentro del medio biológico, la fluorescencia emitida desde la superficie del medio puede variar desde isotrópica hasta fuertemente anisotrópica, dependiendo de si el medio es altamente absorbente, diluido o turbio. Con el fin de explorar el efecto de la absorción y dispersión de la fluorescencia emitida desde la superficie de los tejidos, se han utilizado ampliamente simulaciones de Monte-Carlo. Dichas simulaciones fueron realizadas utilizando perfiles de haces de excitación finitos y complejas geometrías de emisión y excitación para simular la distribución de la luz en medios turbios.[32-34] Los resultados de estas simulaciones han permitido varios hallazgos interesantes, entre los que se destaca que el perfil del haz de excitación es un factor muy importante. Esto es debido a que afecta en gran medida la distribución de la luz de excitación en el tejido.

Por otro lado, el efecto de la absorción, particularmente debido a la hemoglobina, sobre la fluorescencia medida aumenta cuando la zona de medición se aleja de la de iluminación en la superficie del tejido, lo que produce una disminución de la intensidad de la fluorescencia. Además, el efecto de esta absorción depende de la longitud de onda, lo que sugiere que también afecta la forma del espectro de fluorescencia. Esto es evidente cuando se compara la fluorescencia emitida en dos longitudes de onda en las que la hemoglobina tiene diferentes características de absorción. A medida que la distancia de la fibra de recolección aumenta con respecto a la fibra de iluminación, la relación de la intensidad de fluorescencia en las dos longitudes de onda de emisión se desvía de la unidad. Esto sugiere que la recolección de la luz emitida desde el sitio del tejido directamente iluminado por la luz de excitación minimizará la distorsión debida a la absorción de hemoglobina, aunque esto reducirá la profundidad de muestreo para las mediciones de fluorescencia.

1.2.4 Espectroscopia de fluorescencia en el análisis de tejidos biológicos.

Tal como se explicó en la sección 1.2.1, cuando la radiación electromagnética incide sobre las células, éstas absorben radiación y se producen transiciones entre los niveles de energía de algunas moléculas que forman la célula, dependiendo de la longitud de onda de la radiación electromagnética y del tipo de moléculas. Una vez que las moléculas se han excitado, puede ocurrir algún proceso de desexcitación, radiativo o no radiativo. En la espectroscopia óptica de fluorescencia se trabaja con los procesos de desexcitación radiativos y, en particular, en el fluorescente.

La emisión luminiscente producida por tejido irradiado con luz ultravioleta puede ser usada para localizar tumores, a través de la fluorescencia natural del tejido (autofluorescencia) o empleando fluoróforos exógenos tipo HpD por ejemplo (derivados hematoporfíricos).[35] De los fluoróforos endógenos que se listaron en la tabla 1.1, hay

algunos que juegan un rol fundamental en las transformaciones que ocurren en los tejidos durante la carcinogénesis. Ellos son el triptófano, el colágeno, la elastina, el NADH, el FAD y las porfirinas. Sus espectros de emisión y absorción se muestran en las figura 1.13.

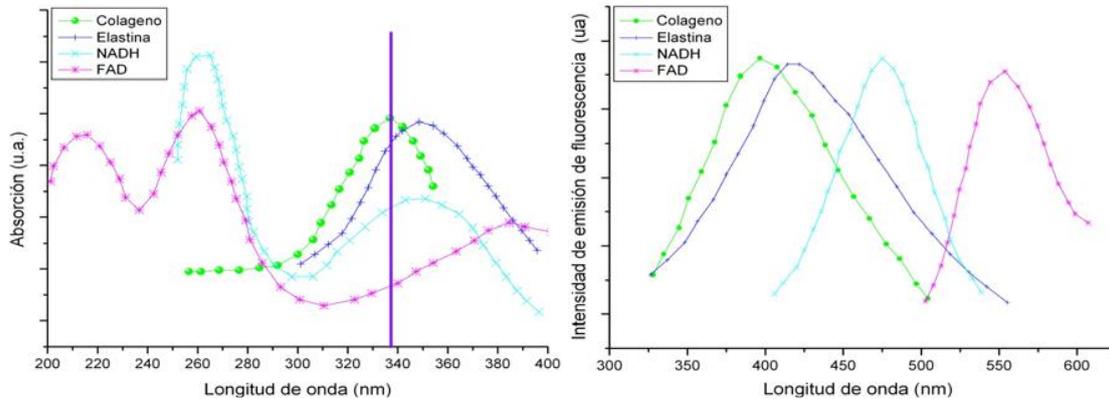


Figura 1.13 espectros de absorción y emisión de los fluoróforos (colágeno, elastina, NADH y FAD. Adaptado de [36].

Dado que mediante la espectroscopia de fluorescencia se puede obtener información significativa acerca de los fluoróforos presentes, como también de su ambiente local, este método se puede aprovechar entonces para detectar zonas de tejidos biológicos que presenten una condición diferente de la “normal”. Tal como se mencionó anteriormente las células afectadas por ciertas enfermedades, y en particular las oncológicas, suelen presentar alteraciones en su metabolismo, y/o en su estructura. Esto repercute en variaciones de los espectros de emisión de fluorescencia de los tejidos formados por células afectadas, respecto de los de células normales. Estas diferencias en la emisión de fluorescencia que se presentan según el estado de las células, dependen de, al menos, uno de los múltiples parámetros que ya fueron mencionados y se describen a continuación: la concentración de fluoróforos y su distribución a lo largo del tejido, el medio que rodea al fluoróforo, la arquitectura particular del tejido y sus propiedades de atenuación de la luz, etc. Además, todas estas características pueden variar significativamente entre capas de tejido, por lo que la espectroscopia de fluorescencia en tejidos también se ve influenciada por las características de las diferentes subcapas.

La variación de la concentración de fluoróforos dentro de los tejidos, puede atribuirse a diversas causas. Por ejemplo, existen varias sustancias endógenas que no se pueden producir en un tejido enfermo, o por el contrario se producen en exceso. En otros casos, la distribución o forma del fluoróforo puede variar a lo largo del tejido. Por ejemplo, NADH es una molécula extremadamente fluorescente en su forma reducida, pero no es fluorescente en la forma oxidada. Por lo tanto, dependiendo de la fase específica de una

reacción metabólica particular que pueda estar ocurriendo, la fracción relativa de NADH en las formas reducidas y oxidadas puede variar enormemente. [37]

Otra característica que muestra diferencias en la emisión de fluorescencia de varios tipos de tejidos es la diferencia en el entorno local que rodea al fluoróforo. El cambio en la microestructura (entorno local) que rodea a un fluoróforo particular puede tener efectos significativos en sus propiedades de fluorescencia. Las propiedades de fluorescencia que en general se ven afectadas por tales cambios incluyen el rendimiento cuántico de fluorescencia, el ancho de banda de su máximo espectral de emisión y su tiempo de vida de fluorescencia. Todos estos factores pueden tener un impacto significativo en la emisión global de fluorescencia de toda la muestra de tejido. Además del entorno local que rodea al fluoróforo, la estructura del tejido a mayor escala también desempeña un papel importante en la fluorescencia resultante debido a sus propiedades ópticas específicas (por ejemplo, índice de refracción, secciones transversales de dispersión, etc.).

Otro factor que a menudo juega un papel importante en las diferencias asociadas con diversos estados patológicos del tejido es la presencia de cromóforos no fluorescentes (que absorben la luz pero no emiten fluorescencia), como la hemoglobina, por ejemplo. Estos cromóforos, que pueden variar en concentración dependiendo del estado del tejido, pueden absorber varias longitudes de onda de la luz fluorescente emitida por el tejido o incluso absorber longitudes de onda específicas de la luz de excitación utilizada para el análisis. Un ejemplo de este caso se observa en el diagnóstico de fluorescencia del tejido tumoral, que típicamente tiene un aumento en la vascularización generada por el proceso de angiogénesis. Debido a este aumento de la vascularización, el aumento de la presencia de hemoglobina a menudo provoca un aumento de la absorción de la luz en la región visible del espectro electromagnético.

En resumen, la espectroscopia óptica de fluorescencia, proporciona características discriminantes para estudiar de los cambios bioquímicos y estructurales de los tejidos en los estados de normalidad y patología. Por lo tanto, esta técnica, puede contribuir al diagnóstico médico a través de la observación *in-vivo* de pacientes. Sus ventajas principales se centran en evitar la posible diseminación de células malignas por la biopsia convencional, y evitar los retardos propios de la biopsia convencional en los casos que el diagnóstico rápido de malignidad propio de la biopsia óptica permite un tratamiento inmediato [38-40].

1.2.5 Espectroscopia de fluorescencia en el análisis de fluidos biológicos

El término “cáncer” se refiere a cualquiera de una gran cantidad de enfermedades caracterizadas por el desarrollo de células anormales que se dividen de manera no

controlada y tienen la capacidad de infiltrarse y destruir el tejido corporal normal. Es a menudo una enfermedad mortal y se le atribuye la segunda causa de muerte en el mundo.

La medicina moderna ha desarrollado muchos modos de tratar el cáncer, incluida la extirpación quirúrgica de tumores, la radioterapia, la quimioterapia, la terapia inmunológica, etc. Sin embargo, la clave para un tratamiento eficaz sigue siendo su detección temprana. En cuanto a este punto, cabe destacar que la principal dificultad para su detección radica en que el cáncer no posee síntomas por sí mismo, por lo que suele detectarse cuando avanzó lo suficiente para afectar el funcionamiento de algún órgano, estando ya en una etapa avanzada. Además, los signos y síntomas generales asociados con el cáncer no son específicos de esta enfermedad y varían según la parte del cuerpo afectada.

En la actualidad existe una amplia variedad de métodos de diagnóstico para detectar la presencia de cáncer. Entre ellos pueden mencionarse: la biopsia quirúrgica (o convencional), el examen rectal digital (DRE - Digital Rectal Exam), la tomografía axial computarizada (TC), las imágenes de resonancia magnética (IRM), la prueba de antígeno prostático específico (PSA), las ecografías, exploraciones óseas, tomografías por emisión de positrones (PET), pruebas de médula ósea, cistoscopia, mamografía, etc. Aunque en mayor o menor medida todas estas técnicas son efectivas, cada una de estas pruebas tiene ventajas y desventajas. Por ejemplo algunas de ellas, como la PSA, las pruebas de Papanicolaou y las mamografías, son específicas de determinados órganos, por lo que no sirven para un chequeo general. Otras, como la biopsia convencional, las endoscopías, la biopsia de médula ósea y la cistoscopia, son pruebas invasivas que a menudo causan un cierto malestar al paciente. Por otro lado las exploraciones TC y por MRI, no son tan específicas, son bastante caras y requieren una instrumentación compleja.

De todo lo mencionado anteriormente, se evidencia la necesidad de desarrollar un método simple, relativamente económico y no invasivo para evaluar la presencia de procesos oncológicos en las personas, independientemente de la ubicación del cáncer o del órgano afectado. Además, sería ideal que dicho método se pueda usar tanto para el diagnóstico preliminar en la detección masiva de afectados (screening), como durante el tratamiento, para determinar si la patología ha entrado en remisión. Teniendo en cuenta todas estas consideraciones, es que el presente trabajo de Tesis extendió el estudio de fluorescencia a los fluidos corporales, en particular, a la orina.

La orina es un fluido corporal que se utiliza como método de diagnóstico común de algunas enfermedades, y las pruebas convencionales se basan en análisis fisicoquímicos, organolépticos, moleculares, colorimétricos, bacteriológicos, etc. Entre la gran variedad de productos químicos orgánicos e inorgánicos disueltos en agua que constituyen la orina, se encuentran presentes ciertos fluoróforos endógenos. [41] Debido a estos últimos, la orina posee una fuerte autofluorescencia al ser iluminada con luz de la longitud de onda

apropiada. Según estudios científicos [42-46] los pacientes con cáncer, en cualquier etapa de su enfermedad, presentan variaciones en la concentración de algunas de estas sustancias respecto de las personas sanas, que se expresan claramente en los espectros de fluorescencia de sus orinas. Aunque estos hallazgos apoyan la hipótesis de que la espectroscopia de fluorescencia endógena puede proporcionar un método para medir diversos cambios pato-bioquímicos en la orina, y de esta manera lograr el diagnóstico de la patología, la investigación en este campo al momento de comenzar este trabajo se encontraba en la fase inicial. Además, el estudio sobre la espectroscopia óptica de orina en oncología diagnóstica todavía no se ha optimizado ni experimental ni estadísticamente.

En base a todos estos antecedentes, el propósito del trabajo que se presenta en la parte 2 es determinar si las características espectrales de emisión de fluorescencia endógena de la orina proporcionan una diferencia estadísticamente significativa entre los pacientes con cáncer y los sujetos normales. Para ello se emplearon diferentes metodologías, basadas en el álgebra matricial y frecuentemente empleadas para el análisis de datos espectroscópicos, para la construcción de modelos “blandos” que faciliten la interpretación de datos multivariados y permitan el desarrollo de herramientas para el diagnóstico de situaciones patológicas.

Parte I

Aplicación de técnicas ópticas sobre tejidos para el diagnóstico no invasivo de enfermedades

Capítulo 2 Materiales y métodos empleados para el estudio de tejidos

2.1 Espectroscopia de fluorescencia sobre tejidos

En la figura 2.1 se muestra un esquema de los componentes básicos del instrumento utilizado para la espectroscopia de fluorescencia de tejidos en el presente trabajo de Tesis. Consiste en una fuente de luz, un conducto flexible que contiene fibras ópticas para la iluminación y la recolección de luz, una red de difracción que separa la luz emitida en sus respectivas longitudes de onda, y un detector que mide la intensidad en estas longitudes de onda. De esta manera, la emisión de fluorescencia, es medida en una configuración geométrica en la cual la iluminación y recolección de la luz corresponden a la misma superficie del tejido biológico. Los diversos tipos de instrumentos empleados para estas mediciones tienen esencialmente los mismos componentes básicos.

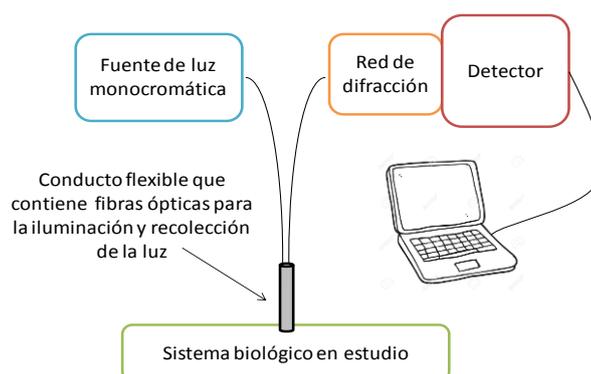


Figura 2.1 Esquema de los componentes básicos de un instrumento usado para espectroscopia óptica de fluorescencia de tejidos.

En lo que a la fuente de luz se refiere, generalmente se utilizan los láseres como fuentes de luz de excitación monocromática, por poseer una salida espectral muy estrecha. Las lámparas también pueden usarse como fuentes “cuasimonocromática” cuando se acoplan con un monocromador o filtro de banda estrecha. Los criterios principales para la selección de fuentes de luz, además del ancho de banda, son la portabilidad, la potencia de salida, la sintonización de longitud de onda, la eficiencia de acoplamiento (en fibras ópticas) y la necesidad de fuentes de luz de onda pulsada versus continua.

Para realizar las medidas de fluorescencia de tejidos del presente estudio se utilizó un espectroflurómetro, que emite luz de 405 nm y tiene la capacidad de sensar la emisión

fluorescente en el rango de 500 a 800 nm. Su cabezal de medición posee un área sensible de 200 μm de diámetro. La figura 2.2 es una foto del cabezal de medición.

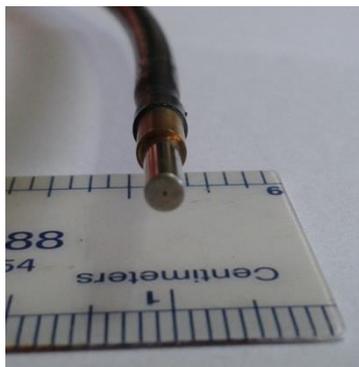


Figura 2.2 Fotografía del cabezal de medición del espectrofluorómetro utilizado.

Para la calibración del espectrofluorómetro, se utilizó un dispositivo provisto conjuntamente con el equipo. Este dispositivo de calibración posee dos orificios, uno de los cuales tiene un vidrio fluorescente en su interior (FG: fluor glass), y el otro un blanco absorbente que no emite señal (BG: background). Previamente a realizar medidas con el espectrofluorómetro, es necesario calibrarlo, insertando el cabezal de medición alternativamente en los dos orificios, de manera de asignar el 0 y el 100 de la escala (unidades arbitrarias) a las medidas del BG y del FG respectivamente.

Los parámetros de medición que se deben seleccionar en el software del espectrofluorómetro son: las longitudes de onda de comienzo y fin del espectro (en el rango 400 a 800 nm), el tiempo de integración y la cantidad de escaneos promedio. Para todas las configuraciones el intervalo entre longitudes de onda sucesivas de emisión es de 1 nm (paso fijo). El tiempo de integración (IT por su sigla en inglés) es la cantidad de milisegundos empleados para detectar cuentas cada longitud de onda de emisión. En este equipo se pueden elegir valores entre 10 y 60000 ms. La cantidad de escaneos promedio, (AV - average scan) es el número de espectros completos que el equipo promedia, para dar el espectro resultante. Puede optarse desde un escaneo simple (AV=1) hasta 32 escaneos promediados.

Con el fin de poner a punto la técnica se realizaron medidas de prueba con este espectrofluorómetro. Para ello se realizaron reiteradas calibraciones y se estudió el comportamiento del equipo al modificar el tiempo de integración y la cantidad de escaneos promedio.

Para analizar la influencia del tiempo de integración en los espectros, se realizó una serie de medidas del vidrio fluorescente (FG) del dispositivo de calibración, manteniendo los otros parámetros fijos y cubriendo el rango de uso de los IT. Algunas de las curvas obtenidas se muestran en la figura 2.3, donde puede observarse que al variar el IT la forma del espectro se mantiene, pero varía el ruido de la señal, disminuyendo con el aumento

del IT. De los resultados obtenidos se concluyó que tiempos de integración mayores a los 100 ms resultan apropiados para realizar medidas con una buena relación señal/ruido. Luego, se decidió utilizar IT= 500 ms para todas las medidas que se presentarán en este trabajo de Tesis.

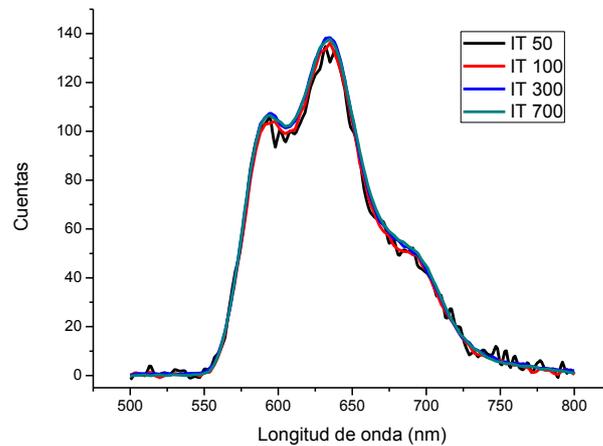


Figura 2.3 Espectros de emisión del vidrio fluorescente (FG) del dispositivo de calibración, variando el IT en todo su rango de uso.

Una vez definido el IT, se realizaron pruebas variando la cantidad de escaneos que se promedian (AV). Para ello se usó nuevamente el vidrio fluorescente del kit de calibración, y se varió todo el rango posible de AV (entre 1 y 32). Tal como puede observarse en la figura 2.4, los espectros obtenidos no presentan diferencias significativas. Esto muestra que el equipo posee una excelente respuesta aún para escaneos simples (AV=1), a este tiempo de integración.

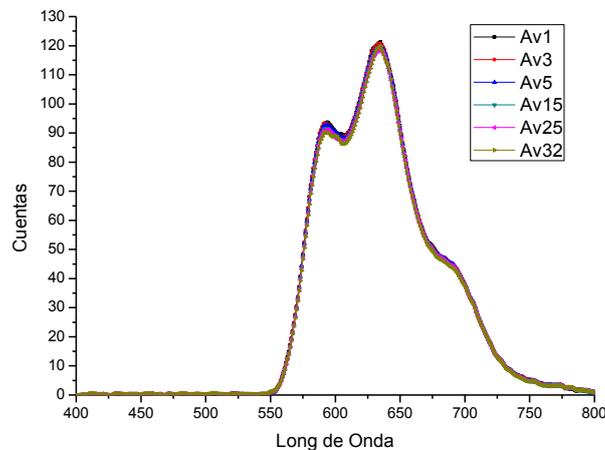


Figura 2.4 Espectros de emisión del vidrio fluorescente (FG) del dispositivo de calibración, variando la cantidad de escaneos promediados, con IT= 500.

Luego de realizar estas comprobaciones, y verificar el comportamiento del espectrofluorómetro, se resolvió utilizar este equipo con $IT=500$ y $Av=1$ para medir todos los espectros de emisión de fluorescencia (EEm) tanto natural como inducida, de tejidos que se presentan en este trabajo de Tesis.

2.2 Diagnóstico fotodinámico

El diagnóstico fotodinámico (DFD), que también es conocido por sus siglas en inglés como PDD (Photodynamic diagnosis), se basa en la identificación de tejidos patológicos, mediante la detección de la fluorescencia de sustancias se localizan selectivamente en ellos. Estas sustancias pueden ser tanto fluoróforos endógenos, como fotosensibilizadores (FS) exógenos. Los FS pueden ser administrados por vía sistémica (mediante una inyección) o tópica (en forma de crema) y, luego de esperar cierto tiempo, se elimina de las células normales quedando acumulado en las patológicas. El tiempo para lograr la mayor diferencia en la concentración de FS entre ambos tejidos es particular para cada FS, pero varía entre unas pocas horas y tres días. Una vez transcurrido ese tiempo, se ilumina la zona a estudiar con radiación de la longitud de onda apropiada según el FS (siempre en el rango UV). Al absorber esta radiación, los FS emiten fluorescencia roja (en el rango espectral visible). De esta manera, se puede delimitar el área de extensión de la lesión mediante la simple observación de las zonas que fluorescen.

Dada la selectividad del método, con el DFD es muy frecuente detectar lesiones “extratumorales” que se localizan en áreas que a simple vista parecen normales. En la terapia fotodinámica (que se explica en la siguiente sección), el DFD es sumamente útil y se usa para definir el área a tratar, el margen de seguridad y la dosis de luz a administrar en cada tratamiento en particular.

Si bien al iluminar con luz UV la fluorescencia de los tejidos que acumularon el FS se observa a simple vista, se pueden también registrar los espectros de emisión de fluorescencia para estimar la cantidad de droga acumulada en la lesión. [47] Esto es particularmente útil en la TFD, ya que sirve para medir de manera indirecta el efecto global del tratamiento. Tal como se explicará más adelante, esto se logra comparando los EEm que se realizan antes y luego de la iluminación.

2.3 Terapia fotodinámica

La aplicación de fotosensibilizadores, utilizados en forma controlada y dirigida sobre determinados tumores, ha dado lugar a la terapia fotodinámica (TFD) del cáncer. [48] Esta se fundamenta en el efecto citotóxico inducido en el tejido tumoral por la acción

combinada de luz, oxígeno y una droga (el fotosensibilizador), que se acumula con un cierto grado de selectividad en dicho tejido.

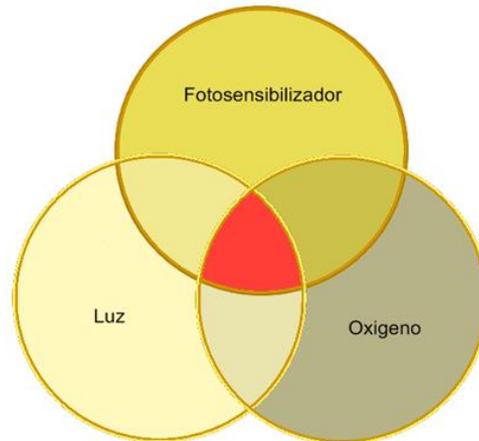


Figura 2.5 Esquema de los tres elementos fundamentales de la TFD.

Una sustancia se comporta como agente fotosensibilizador (FS) cuando al ser irradiada por luz ultravioleta (UV) o visible, se “activa” y genera especies reactivas de oxígeno (1O_2 y otras especies reactivas del oxígeno, ROS), que desencadenan la oxidación violenta de los componentes celulares. Este fenómeno deriva en una serie de procesos que conducen a una suerte de “implosión celular” que destruye las células cancerosas. La figura 2.6 describe esquemáticamente el procedimiento completo, tanto para la TFD, como para el DFD.

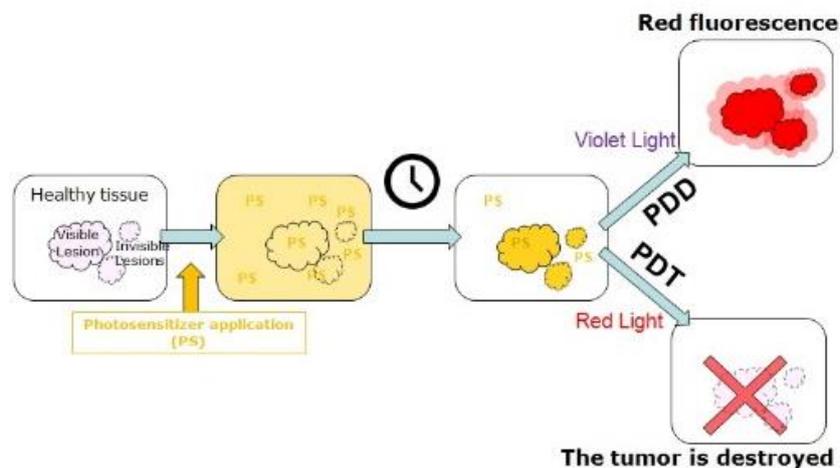


Figura 2.6 Esquema representativo del DFD y la TFD.

Del mismo modo que en el caso del cáncer, cuando son administradas en un organismo vivo en forma controlada estas sustancias se convierten en una herramienta muy eficaz para el tratamiento de numerosas patologías, ya que al ser irradiadas actúan como agentes antibacterianos, antifúngicos, antivirales y anticancerígenos.

La TFD tiene uso clínico en el tratamiento de muchas enfermedades. Está aprobada para diversas aplicaciones por las Agencias de Salud de Europa, de los Estados Unidos de América, de Canadá y de Japón. Allí, es el método elegido para el tratamiento de la degeneración macular y es de relevancia contra diferentes tipos de cáncer (de piel, esófago, cabeza y cuello, pulmón, y vejiga) debido a que resulta una terapia mínimamente invasiva, con alta eficiencia curativa, carente de efectos colaterales significativos y posee muy buenos resultados estéticos.

Esta terapia es especialmente indicada para tratamiento de zonas del cuerpo fácilmente accesibles, como son las mucosas y epitelios superficiales, y por eso se considera un tratamiento no invasivo. Sin embargo, también da la posibilidad extender su aplicación a zonas de difícil acceso, a través de métodos invasivos que involucran punciones, como son el caso del pulmón y la parótida que se presentan en el Capítulo 3.

Tal como se mencionó anteriormente, además de la presencia de oxígeno en el lugar, existen otros dos factores claves a tener en cuenta para el éxito de la TFD: el fotosensibilizador aplicado y la luz con que se ilumina. En lo que resta de esta sección, se presenta una descripción detallada de los FS y fuentes de luz empleados en el presente trabajo.

Los principales fotosensibilizadores empleados en la TFD son los derivados de porfirina y del ácido 5-amino levulínico (ALA). En el presente trabajo utilizamos m-tetrahydroxyfenil clorina (mTHPC, Temoporfin o Foscan®), para la administración sistémica y ALA (ácido 5-amino levulínico) para uso tópico.

Estructura y función de la mTHPC.

La m-THPC es un eficiente generador fotoquímico de oxígeno molecular en estado singulete. Su estructura molecular se muestra en la figura 2.7, junto con su espectro de absorción. Tal como puede observarse, este último presenta un pico característico a 652 nm y la banda de Soret a 400 nm.

Entre las principales ventajas de este FS por sobre otros, se destaca que se requieren dosis bajas para producir el efecto fotodinámico, posee una intensa banda de absorción en la región del rojo del espectro visible y posee un comportamiento farmacocinético favorable, que aumenta la selectividad del tejido tumoral sobre el tejido normal. Además la m-THPC tiene baja o nula toxicidad en la oscuridad.

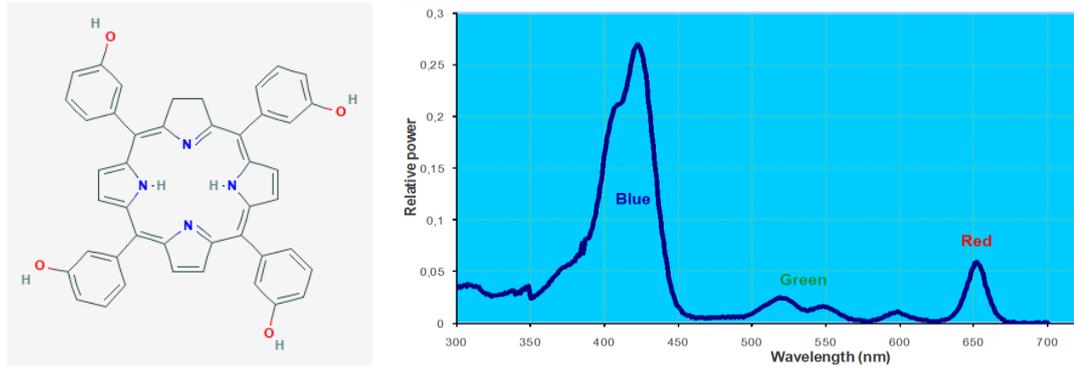


Figura 2.7 Estructura [49] y espectro de absorción de la m-THPC (Temoporfirina).

Estructura y función del ALA

Tal como se explicó en la sección 1.2.2, el ácido 5-aminolevulínico (5-ALA) es un precursor natural de la protoporfirina IX (PpIX). El 5-ALA es captado particularmente por las células cancerosas en las que se metaboliza a protoporfirina IX (PpIX), compuesto fotoactivo que se acumula dentro de las células lesionadas en relación con el tejido normal. La figura 2.8 muestra la estructura de ambos compuestos.

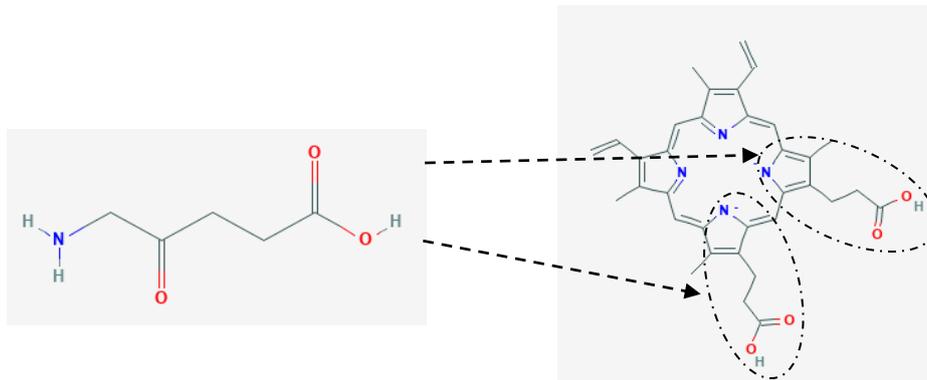


Figura 2.8 Representación esquemática de la incorporación del ALA en la biosíntesis de la PpIX. [49]

En presencia de luz visible, la fluorescencia de la PpIX en ciertos tejidos puede ser utilizada para diagnóstico fotodinámico (DFD), o bien como tratamiento (TFD), causando daños en los componentes celulares y destruyendo a las células diana. Este último es especialmente utilizado en terapia de cáncer de piel

La PpIX se caracteriza por un máximo de absorción a 405 nm y máximos de fluorescencia a 635 y 705 nm.

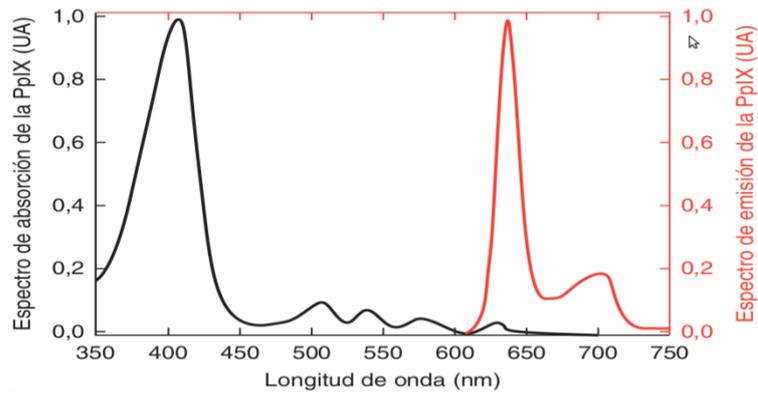


Figura 2.9 Espectros de absorción (en negro) y de emisión (en rojo) de la PpIX. Adaptado de [50].

Por otra parte, más allá de las características propias de la fuente de luz que se utiliza, y su relación con las bandas de absorción del fotosensibilizador, tanto para el DFD como para la TFD, es muy importante tener en cuenta la penetración de la luz en el tejido, tal como se explicó previamente en la sección 1.2.3. El rango de longitudes de onda en el cual la penetración de la luz es máxima en determinado tejido, se denomina “ventana óptica”. En la figura 2.10 se muestra la ventana óptica para el caso de la piel humana. Tener en cuenta esta información combinada con la absorción de la droga fotoactivable y la oxigenación del tejido es fundamental para lograr el éxito de la terapia.

Además de la ventana óptica, es importante tener en cuenta que la radiación electromagnética de longitudes de onda por debajo de los 400 nm tiene la capacidad de inducir fotólisis directa en determinadas moléculas, en ausencia de fotosensibilizador. Por ello, es fundamental destacar que la radiación electromagnética que se utiliza en las aplicaciones de la TFD, y en todos los casos presentados en el capítulo 3, se encuentra por encima de ese valor, lo que garantiza que la absorción de esa radiación no produzca alteraciones químicas en ninguna de las moléculas de importancia biológica.

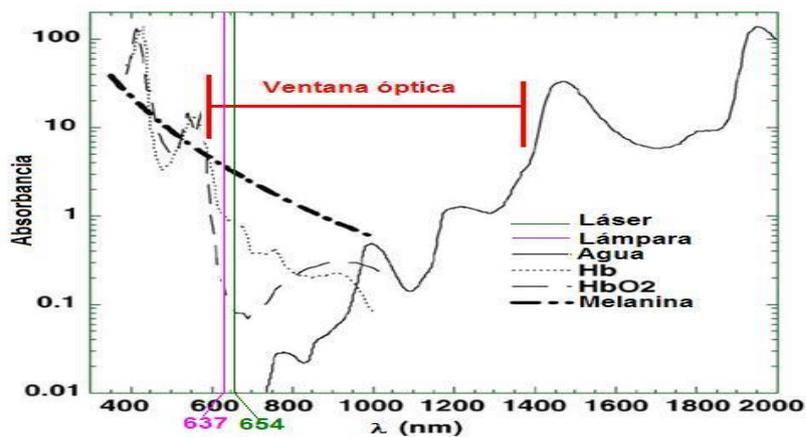


Figura 2.10 ventana óptica de la piel humana. Sobre ésta se ven superpuestos los máximos de emisión de dos fuentes de luz que se utilizaron en el presente trabajo y se describen a continuación en el texto: láser de 654nm (en verde), y la lámpara led de 637nm (en violeta).

Adaptado de [51]

La luz azul penetra menos eficientemente a través del tejido, mientras que las radiaciones roja e infrarroja penetran más profundamente (figura 1.11). La región entre 600 y 1200 nm a menudo se denomina ventana óptica de tejido. Sin embargo, la luz de longitudes de onda mayor a 800 nm no tiene la suficiente energía como generar $^1\text{O}_2$, de manera de iniciar una reacción fotodinámica. No existe una fuente de luz "ideal" para todas las indicaciones de TFD, incluso con el mismo FS. Por lo tanto, para la elección de la fuente hay que tener en cuenta las características espectroscópicas del FS (espectros de excitación, y emisión de fluorescencia), la patología en particular (ubicación, tamaño de las lesiones, accesibilidad y características del tejido), el costo y el tamaño del equipo. La eficacia clínica de la PDT depende en forma compleja de la dosimetría: dosis total de luz, tiempo de exposición a la luz y modo de suministro de luz (simple o fraccionado). La elección de combinaciones óptimas de FS, fuentes de luz y parámetros de tratamiento es crucial para una PDT exitosa.

Para activar el FS en los casos de TFD que se presentan en el Capítulo 3 se utilizaron dos fuentes de luz: un láser de diodo y una lámpara LED. El láser de diodo emite en 654 nm, con un ancho de banda de 2,4 nm y su potencia de emisión es variable, pudiendo regularse hasta un máximo de 2 W (limitado según la fibra óptica a utilizar). En algunos casos, el láser se acopló a una fibra óptica cuyo extremo libre tiene una microlente que hace que el haz diverja, de manera de lograr un área de iluminación amplia y circular. La lámpara LED que se utilizó en el caso 7 (Cap. 3) fue completamente caracterizada para los tratamientos de TFD en el marco de la Tesis doctoral de la Lic. Ma. Eugenia Etcheverry. Dicha lámpara posee su pico de emisión en 637 nm, con un ancho de banda de 18,1 nm y tiene una potencia de 1,06 W. En la figura 2.11 se muestra su espectro de emisión, junto con el del láser de diodo y el de absorción de la m-THPC en solución.

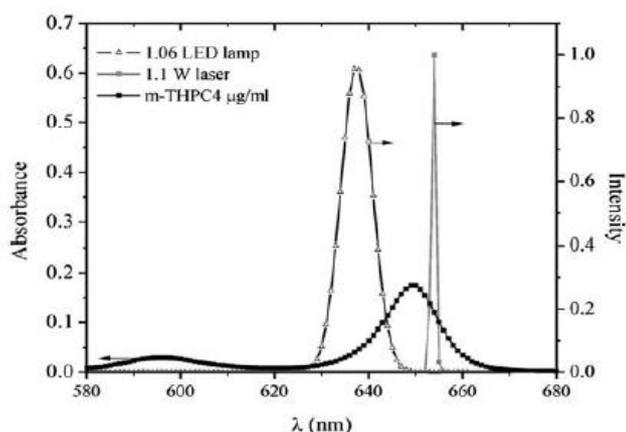


Figura 2.11 Comparación entre láser y lámpara LED con el espectro de absorción de una solución de m-THPC. La escala de intensidad es arbitraria. Adaptado de [51]

Las fuentes de luz mencionadas anteriormente fueron utilizadas en los procedimientos de TFD. Para otros procedimientos que también se presentan en el Capítulo 3 se utilizaron diferentes fuentes, específicas para cada caso, a saber:

1. Para la fotocoagulación y fotodesnaturalización del lecho quirúrgico del caso 5 se utilizó el laser de diodo Diomed, que se muestra en la figura 2.12. Dicho laser emite en 810 nm y tiene una potencia máxima de 30 W.



Figura 2.12 Laser de diodo Diomed 30W.

2. Las imágenes de fluorescencia de los DFD se realizaron con una lupa dermatológica comercial destinada a tal fin, y con una luz LED que emite en 400 nm, según el caso. La lupa, que se muestra en la figura 2.13, contiene de 4 tubos F4T5/BLB, con una potencia total de 16 W (4 W cada uno), que emiten en la región espectral UVA.

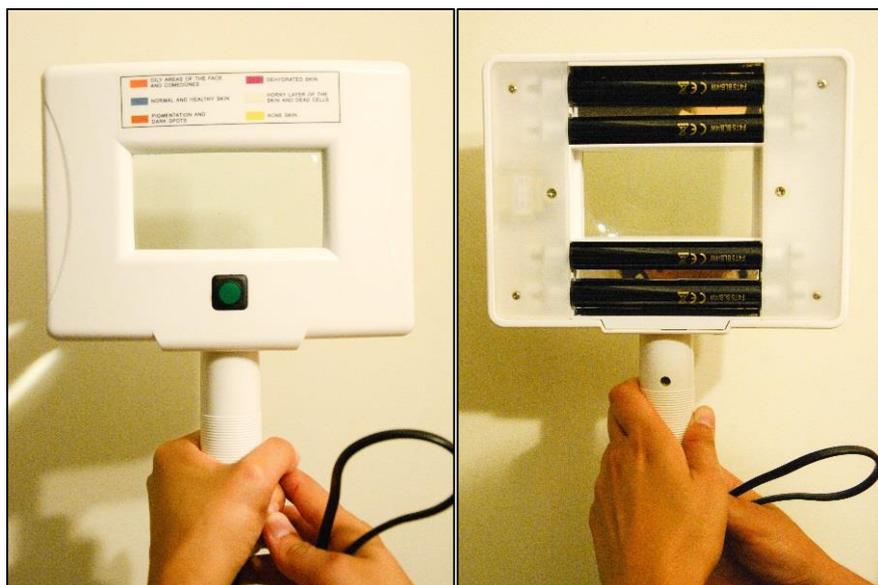


Figura 2.13 Lupa comercial que emite luz en el rango UV.

3. Para la obtención de espectros de autofluorescencia y de fluorescencia inducida de tejidos se utilizó la fuente de luz de espectrofluorómetro que se describió en la sección 2.1, que emite en 405 nm.

En ciertas ocasiones, en los tratamientos de terapia fotodinámica, resulta de interés cuantificar de alguna manera el éxito de la aplicación. Para ello existe una técnica muy sencilla, fue publicada en 1999 por Claes af Klinteberg y colegas [52], y se describe a continuación. Esta técnica estima el resultado de la terapia a través de una medida indirecta, monitoreando la cantidad de fotoproductos que se generan *in vivo* por la TFD.

Luego de la aplicación de FS, y de esperado el tiempo necesario para la localización del mismo en la zona a tratar, se registra un espectro de emisión de fluorescencia (Figura 2.14 A). Este primer espectro se realiza en un momento previo, muy cercano a la activación del fármaco. Luego, se realiza el tratamiento, e inmediatamente después, se obtiene un segundo espectro de fluorescencia en el mismo lugar (Figura 2.14 B). La señal fluorescente debida al FS se designa A_0 para mediciones antes de la irradiación, y A_1 para mediciones después de la irradiación, mientras que C_0 representa la intensidad de fluorescencia en la región de los fotoproductos antes de la irradiación y C_1 la intensidad de fluorescencia de los fotoproductos después de la irradiación.

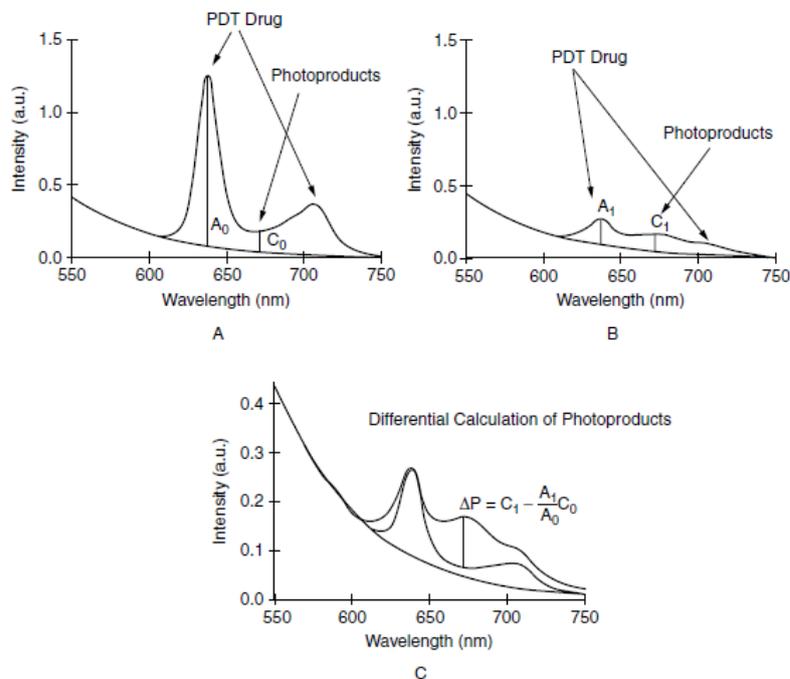


Figura 2.14 Método de fluorescencia para controlar la producción de fotoproductos TFD producidos *in vivo* durante el tratamiento. Adaptado de [52]

Notar que en todos los casos, la intensidad de fluorescencia se calcula a partir de los espectros corregidos, resultantes de restar la autofluorescencia (que en las figuras 2.14 A-C se muestra representada como la línea base), a los espectros medidos. En la región de longitudes de onda entre 600 y 750 nm la autofluorescencia se calcula ajustando una curva exponencial a cada espectro registrado, utilizando las intensidades entre 530 y 580 nm y entre 750 y 800 nm en el ajuste. Es importante destacar que restar la contribución de la autofluorescencia calculada a los espectros registrados, permite evaluar las intensidades de fluorescencia relacionadas únicamente con PpIX y los fotoproductos. De esta manera, para controlar la cinética de la acumulación de PpIX se calcula la intensidad de fluorescencia libre de fondo (A_0 y A_1) a 635 nm. Por otra parte, para evaluar la formación de fotoproductos (ΔP), se usa la intensidad de fluorescencia a 670 nm, después de restar las contribuciones de la autofluorescencia del tejido y la fluorescencia PpIX, respectivamente (Figura 2.14 C).

El aumento de la fluorescencia relacionada con la formación de fotoproductos durante la iluminación terapéutica (ΔP) se evalúa utilizando la ecuación

$$\Delta P = C_1 - \frac{A_1}{A_0} C_0 \quad (2.1)$$

La relación A_1/A_0 , es decir, la relación de la intensidad de fluorescencia de la PpIX pos y pre TFD, es una medida del foto-blanqueamiento (destrucción fotoquímica de un fluoróforo) de la PpIX. Al multiplicar la intensidad de fluorescencia a 670 nm (C_0) por este factor en la ecuación 2.1, la contribución de PpIX puede sustraerse, dando como resultado que el cambio de intensidad calculado sea una buena estimación de la contribución debida únicamente a los fotoproductos generados en la TFD.

2.4 [Termografía IR](#)

La termografía infrarroja es una técnica de examinación del cuerpo humano completamente no invasiva, segura y efectiva. Es útil y eficaz para el diagnóstico de muchas patologías entre las cuales figuran cáncer de mama, trastornos del sistema nervioso y del metabolismo, artritis, trastornos vasculares, alergias, quemaduras, heridas, melanomas y lesiones en tejidos blandos.

Las imágenes de termografía IR permiten detectar y medir pequeños incrementos de temperatura a nivel superficial (piel), que pueden ser originados por diversas causas. Un ejemplo es el aumento de la vascularización en lesiones neoplásicas y preneoplásicas. Dicha información es de suma importancia en las aplicaciones de terapia fotodinámica del cáncer, ya que para el éxito del tratamiento son fundamentales la llegada de oxígeno y de la droga fototactivable al tumor. En este caso la termografía IR se usa con el fin de verificar la presencia de neovascularización de las áreas tumorales.

La figura 2.15 es una imagen de la cámara utilizada (Thermovision® 550, de AGEMA), y en la tabla 2.1 se detallan sus características técnicas.



Figura 2.15 Cámara IR utilizada.

Rango de medida	-20°C a 250° Hasta 1500°C con filtro
Exactitud	± 2% del rango o ± 2°C
Sensibilidad	<0,1°C
Detector	320x240 pixeles
Campo de visión	20° x 15° (alto x ancho)
Lentes	Germanio
Tamaño	220 x 132 x 140 mm
Peso	2 Kg

Tabla 2.1: Características técnicas de la cámara IR AGEMA Thermovision® 550.

Capítulo 3 . Biopsias ópticas de tejidos

A lo largo del presente estudio se han realizado observaciones ópticas “*in vivo*” en pacientes afectados por diferentes tipos de lesiones: verrugas vulgares, neoplasias cutáneas, lesiones por HPV en dermis y en mucosas, etc. Dichos pacientes fueron seleccionados por el médico oncólogo Dr. Horacio Poteca, quien es especialista en Terapia Fotodinámica del Cáncer. En el presente Capítulo se presentan algunos estudios de casos. Con el fin de obtener resultados que puedan contribuir al diagnóstico clínico se analizaron los espectros de emisión de fluorescencia de las zonas afectadas y de regiones sanas adyacentes. Se observaron tanto los espectros de fluorescencia natural (fluoróforos endógenos), así como también los de la fluorescencia inducida en las partes afectadas luego de la aplicación de las drogas fotoactivables que se utilizan en los tratamientos de TFD y que actúan selectivamente en las células tumorales. Durante las sesiones de TFD se tomaron y analizaron imágenes de fluorescencia para delimitar las áreas a tratar, y en algunos casos se realizaron observaciones complementarias por termografía infrarroja de las mismas áreas. Tal como se explicó en el Capítulo 2 las imágenes termográficas se utilizan tanto como técnica de diagnóstico complementario como para el control del incremento de la temperatura durante la TFD. A continuación se describen los 9 casos seleccionados:

Caso 1. Espectros típicos de piel sana

Se tomaron espectros de auto fluorescencia de tejido sano, en particular, en este caso se muestran los obtenidos sobre diferentes partes de una mano. El objetivo fue obtener espectros de referencia y analizar si estos varían en función de las diferentes estructuras presentes en profundidad. Para ello se midió la fluorescencia de la delgada piel de la mano, en lugares sobre las venas superficiales (vena dorsal y ramificaciones puntos 4, 5 y 6), sobre los nudillos (puntos 2, 3 y 7), y en el dorso mismo (punto 1). Entre las mediciones realizadas sobre los nudillos, en una de ellas se hizo poniendo especial énfasis en posicionar la fibra del fluorómetro sobre el tendón del músculo extensor del dedo índice (punto 7).

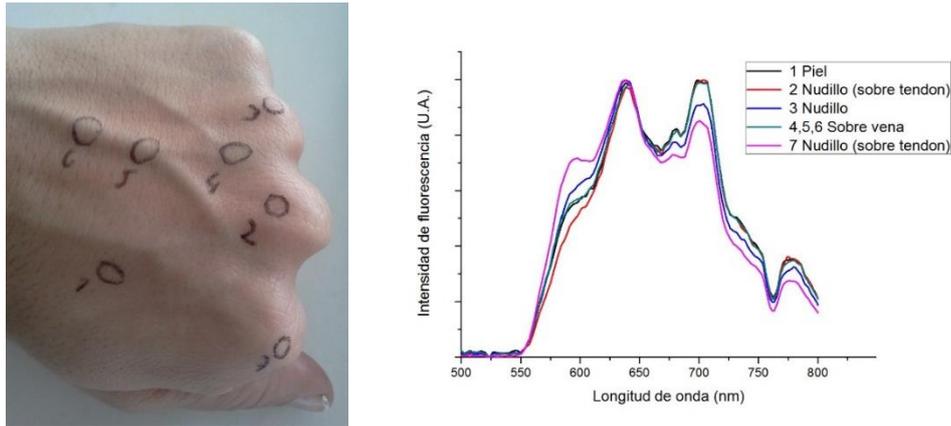


Figura 3.1 Ubicación de los puntos muestreados (Izq.) y espectros obtenidos (Der.).

Según puede observarse en la figura 3.1, todos los espectros presentan la misma estructura, y sólo se diferencian levemente en la intensidad relativa de algunos picos.

Todos los espectros presentan bandas dominantes en 637 y 700 nm, picos secundarios en 670 y 775 nm, y un hombro en 595 nm. Dado que este último hombro es mucho más pronunciado para el espectro del punto 7 que para el resto, posiblemente esté relacionado con la presencia del tendón. Por otro lado, los espectros 1, 4, 5 y 6 no evidencian ninguna diferencia, indicando que no son afectados por la presencia en profundidad de las venas.

Para ilustrar las diferencias en los espectros dadas por los diferentes tejidos, en la figura 3.2 se muestra un espectro correspondiente a epitelio sano de cuello de útero tomado *in-vivo*, uno de tejido perianal y uno de la palma de la mano.

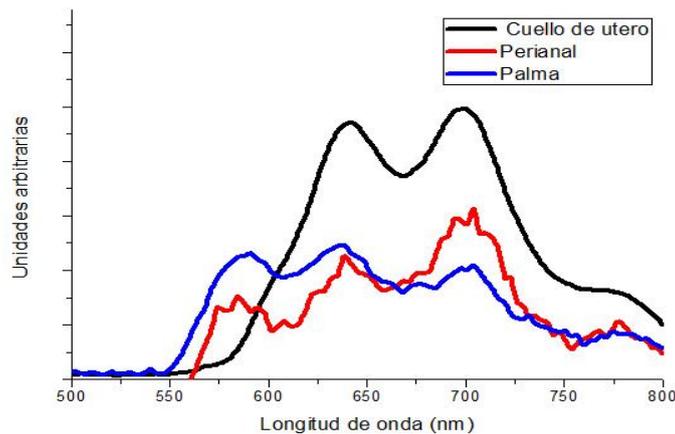


Figura 3.2 Espectros típicos de diferentes tejidos.

Dicha imagen representativa, muestra cómo en cada caso la estructura de los registros da cuenta de la variabilidad y cantidad de fluoróforos que pueden contribuir a la fluorescencia de cada tejido, siendo que todos fueron tomados con el mismo instrumento.

Caso 2. Verrugas comunes o vulgares

El primer caso patológico a analizar es el de observaciones por espectroscopia de verrugas comunes o vulgares. Las verrugas son infecciones de la piel y/o las mucosas [53] causadas por ciertos virus pertenecientes a la familia de los papiloma virus humanos (*Papillomaviridae*), ampliamente conocidos como “HPV” por su sigla en inglés (human papillomavirus). Es una infección muy frecuente y que sigue una evolución autorresolutiva en un alto porcentaje de casos, sobre todo en personas no inmunodeprimidas.

Las verrugas se clasifican según su aspecto clínico y según su localización en la piel o en las mucosas (tabla 3.1).

Verrugas cutáneas	Verrugas comunes
	Verrugas planas
	Verrugas palmoplantares
	Verrugas periungueales
	Verrugas filiformes
	Verrugas en mosaico
Verrugas en mucosas	Condilomas acuminados
	Papulosis bowenoide

Tabla 3.1 Clasificación clínica de las verrugas.

La denominada “verruca común” se caracteriza por su superficie rasposa y color grisáceo, además de su protuberancia en forma redondeada u ovalada, con bordes bien definidos. Pueden aparecer en cualquier área cutánea, aunque las manos constituyen la zona más común. Habitualmente las lesiones son múltiples o de gran tamaño, sobre todo en pacientes que padecen inmunodeficiencias. La figura 3.3 presenta una fotografía de la mano del paciente donde pueden observarse las verrugas estudiadas.

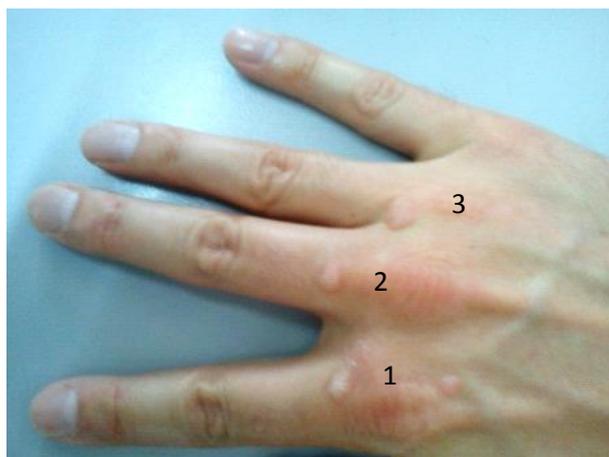


Figura 3.3 Mano del paciente donde se observan las tres verrugas estudiadas en el caso 2.

Se tomaron los espectros de fluorescencia natural de las tres verrugas indicadas en la figura, y de zonas de piel sana cercanas a cada una de ellas, más otra en el dorso de la mano para control. La Tabla 3.2 describe las zonas de la mano en la que se registraron los espectros que se muestran en la figura 3.4.

Verrugas	Blancos (no mostrados en la figura)
1: Dedo índice	4: Dedo índice
2: Dedo mayor	5: Dedo mayor
3: Dedo anular	6: Dedo Anular
	7: Dorso de la mano

Tabla 3.2

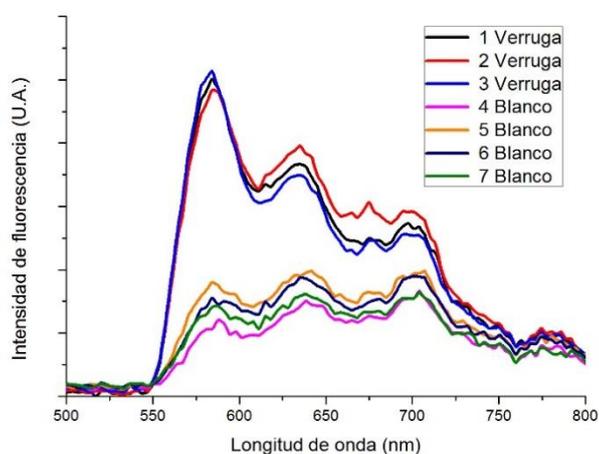


Figura 3.4 Espectros de emisión de fluorescencia del caso 2.

Los espectros medidos en las zonas sin verrugas presentan las mismas bandas (en 595, 635, 670 y 705 y 780 nm) que los del caso 1 y sólo se diferencian levemente en la intensidad relativa de algunas de ellas. También, en la figura 3.4, se observa que existe una clara distinción entre los espectros de las zonas afectadas por las verrugas y los de las zonas sanas. La diferencia más significativa se registró en la intensidad global de fluorescencia, que en todos los casos fue mayor para las zonas afectadas. Además, se observa un notable aumento relativo de la banda de emisión alrededor de los 585 nm, respecto de las otras dos bandas presentes en dichas zonas.

Caso 3. Angioma plano

El siguiente caso corresponde al análisis espectroscópico de un angioma plano ubicado en el área frontal del hombro derecho en la línea axilar. Para identificar la ubicación de los puntos analizados, se utilizó la convención del “reloj”. Dicha configuración es de uso común en medicina, y sirve para ubicar puntos distribuidos en forma circular, utilizando los números en la posición en que se encontrarían en un reloj, con las agujas en el centro de la lesión, o en torno a un punto de interés. Un esquema de dicha configuración, se muestra en la figura 3.5.



Figura 3.5 Esquema de la convención de reloj.

Se midió el espectro de fluorescencia natural en cinco puntos sobre el angioma: en el centro, a las 3, a las 6, a las 9 y a las 12; y en dos puntos fuera del angioma que se consideraron como “blanco”. Las figuras 3.6 y 3.7 muestran una fotografía del angioma y los espectros obtenidos, respectivamente.



Figura 3.6. Fotografía del paciente de caso 3 donde se observa el angioma.

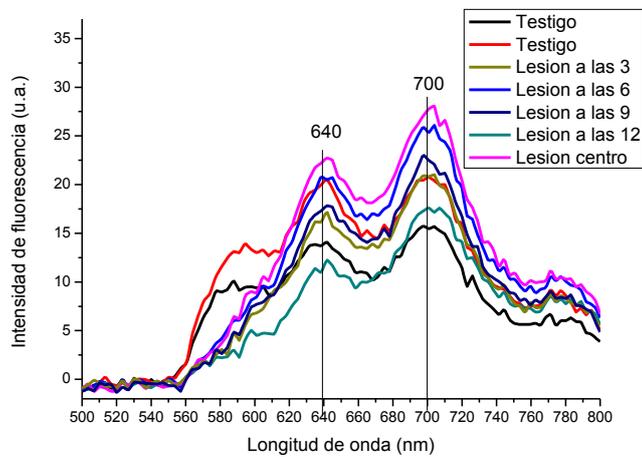


Figura 3.7 Espectros obtenidos. Los números indicados en las curvas son los correspondientes a la ubicación espacial de las lesiones observadas, según la convención del "reloj".

En todos los casos los espectros presentan formas similares y poseen dos bandas dominantes: a 640 nm y a 700 nm. En el caso de los blancos (o testigos), además de esas dos bandas, se observa un incremento de la emisión de fluorescencia alrededor de los 580 nm respecto de las medidas sobre el angioma. Esto puede observarse con mayor claridad en la figura 3.8, donde se presentan los mismos espectros que en la figura 3.7, pero normalizados al máximo de cada uno.

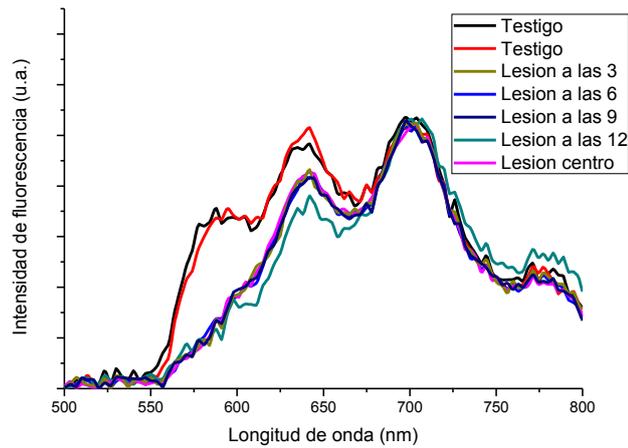


Figura 3.8 Espectros del caso 3 normalizados al pico de máxima intensidad.

En esta otra representación, que implica sólo un cambio de escala para cada espectro, no sólo se observa con mayor detalle esa diferencia, sino que además queda en evidencia una disminución de la emisión relativa en la zona de la banda centrada a 640 nm para la piel afectada, respecto de la piel sana, que de otra manera sería difícil de identificar.

Para analizar la distribución de temperaturas en la zona del angioma se realizó una termografía IR (figura 3.9).

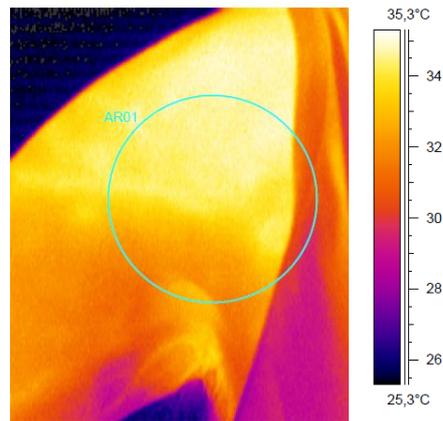


Figura 3.9 Termografía IR de la zona de estudio.

A partir de esta, puede observarse que no existe ninguna diferencia apreciable entre la temperatura del angioma y la de las zonas circundantes.

Caso 4. Proceso inflamatorio crónico liquenoide

El siguiente caso corresponde a una persona de 43 años de edad, que cursaba un proceso inflamatorio crónico liquenoide en el surco vestibular derecho, que luego fue tratado por fotocoagulación con láser de diodo.

El liquen plano oral es una enfermedad inflamatoria crónica que afecta las membranas mucosas dentro de la boca y puede manifestarse como manchas blancas y reticuladas, tejidos rojos e hinchados, o llagas. Estas lesiones causan ardor, dolor u otras molestias similares, que, en general, se pueden controlar. Es una patología que no se contagia por contacto interpersonal, sino que se presenta cuando el sistema inmunitario ataca las células de las membranas mucosas bucales por razones desconocidas (trastorno autoinmunitario). A las personas que tienen esta patología se les recomienda la realización de controles médicos periódicos de la lesión, ya que existe el riesgo asociado de desarrollar cáncer de boca en las zonas afectadas [54].

La figura 3.10 es una fotografía de la boca del paciente, donde puede observarse la lesión.



Figura 3.10 Fotografía donde puede observarse la lesión señalada en amarillo.

En la figura 3.11 se muestran los espectros de fluorescencia natural de la zona afectada y de un testigo fuera de la lesión. Este último fue tomado en las cercanías de la zona afectada, sobre la mucosa sana.

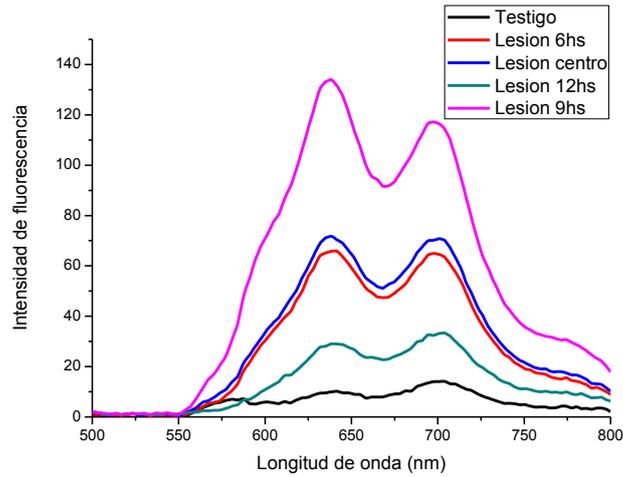


Figura 3.11 Espectros del caso 4, según la convención del reloj.

Si bien los espectros obtenidos presentan una gran variabilidad en la intensidad total de fluorescencia registrada, todos presentan dos bandas principales en 640 y 700 nm. A simple vista parece no haber diferencias significativas entre la zona sana y la afectada. Sin embargo, al normalizar los espectros (figura 3.12) se hace evidente la presencia de un hombro muy pronunciado en la zona de 570-580 nm en el espectro correspondiente a la muestra testigo, que no existe en las zonas afectadas.

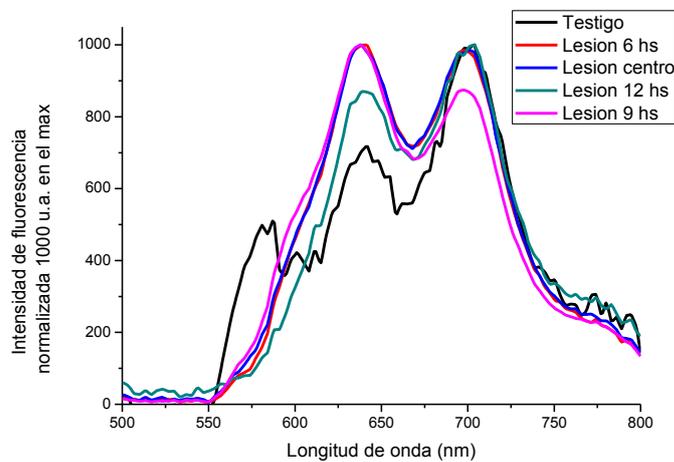


Figura 3.12 Espectros del caso 4 normalizados al pico de máxima intensidad.

Por último, aquí también se observa que la zona afectada presenta mayor intensidad de fluorescencia a los 640 nm respecto del testigo.

Caso 5. Condilomas - VPH6

El siguiente caso analizado es el de un paciente de 46 años de edad, afectado por lesiones perianales tipo condilomas, causadas por el Virus del Papiloma Humano (VPH por sus siglas en Castellano o HPV por sus siglas en Inglés) tipo 6.

Los condilomas, también conocidos como verrugas genitales, son lesiones benignas causadas por la infección del VPH. [55] En la actualidad, ésta es considerada una de las infecciones de transmisión sexual más frecuente y se estima que el 95% de los condilomas son generados por los VPH tipo 6 y 11. Los tipos de VPH causantes de estas lesiones se denominan de “bajo riesgo oncogénico”, es decir que no están tan directamente relacionados con el cáncer. No obstante ello, cerca de una tercera parte de los pacientes con condilomas presentan una co-infección por otros tipos de VPH de alto riesgo oncogénico. Por ello la vacuna cuadrivalente empleada en Argentina extiende la protección contra los tipos VPH-6, VPH-11, VPH-16 y VPH-18, siendo los dos últimos los más importantes entre los de “alto riesgo oncogénico”.

La forma de presentación de los condilomas es muy variable. Frecuentemente se observan verrugas o lesiones sobre elevadas de distinto tamaño (desde pocos milímetros a varios centímetros) con la superficie rugosa, adoptando una forma que recuerda a una coliflor. El número de lesiones también es muy variable, desde pocos elementos (son raras las lesiones únicas) a múltiples lesiones. Dichas lesiones habitualmente afectan las zonas genitales, pero también pueden observarse en la zona perianal, anal o en el periné, y con menor frecuencia pueden afectar otras áreas mucosas como la boca u orofaringe.

En el caso que se presenta el procedimiento se inició con la observación de los espectros de fluorescencia natural de las lesiones. Los espectros registrados se muestran en la figura 3.13.

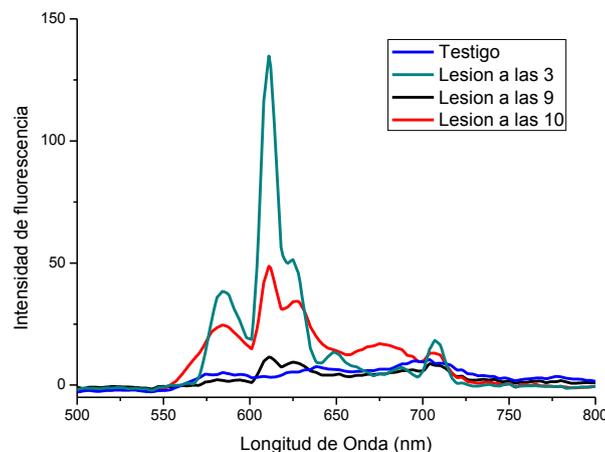


Figura 3.13 Espectros de fluorescencia natural tomados sobre los condilomas y sobre piel sana adyacente (testigo).

Las curvas rotuladas como 3, 9 y 10 corresponden espectros medidos sobre los condilomas, mientras que la muestra indicada como “Testigo” es un espectro control tomado sobre piel sana. Las etiquetas de las curvas se corresponden con la convención del reloj descripta previamente en el caso 3.

De los espectros obtenidos se observa claramente que, a pesar de que las señales de las zonas patológicas se ven afectadas en su intensidad, se conserva la posición de las bandas más importantes y que esa estructura espectral es bien diferente a la que se observa del tejido sano. En este caso en particular, la variación en la intensidad de la señal para las muestras patológicas, se atribuye al efecto del factor angular (coseno del ángulo), debido a la compleja topografía donde las lesiones (condilomas) están localizadas. La banda de máxima intensidad de fluorescencia para todos los espectros medidos sobre los condilomas se ubica a 611 nm. Este es un resultado importante, sobre el que se volverá a discutir más adelante.

Luego de registrar estos espectros, el Dr. Poteca realizó el tratamiento de la zona afectada mediante la fotocoagulación de las lesiones con láser y la fotodesnaturalización del lecho quirúrgico. Esta última se realizó para destruir las células de las capas basales que estaban infectadas por los virus del papiloma. En la figura 3.14 se muestra una fotografía tomada durante este procedimiento.



Figura 3.14 Fotografía tomada durante el tratamiento en la que se observan las lesiones, y el láser en funcionamiento.

Es muy importante mencionar que al realizar este tipo de procedimientos, existe la posibilidad de que los VPH que se encuentran en los tejidos se transformen en fragmentos volátiles, y éstos pueden instalarse con suma facilidad en las células de la piel o mucosas de quienes están presentes. Por ello es de fundamental importancia que el personal interviniente use elementos de protección específicos. En este procedimiento se utilizaron gafas de seguridad, barbijos de triple filtro con filtro viral (3M®) y máscaras. La máscara es

un material transparente que cubre la cara por sobre el barbijo y las gafas, y funciona como deflector del aire que puede llevar fragmentos virales producidos por la acción de fulguración del láser. En la figura 3.15 se muestran las protecciones utilizadas por el personal interviniente en este procedimiento.



Figura 3.15: Fotografía de la tesista donde se pueden observar las protecciones utilizadas.

Como medidas de seguridad adicionales, el lugar contó con un sistema de ventilación forzada (equipo de aire acondicionado), y una toma de absorción del aire desde el lugar de la operación. El aire, previo a ser liberado a los cuatro vientos, atravesó dos filtros de burbujeo, y fue liberado a una altura de 6 m del piso.

Caso 6. Lesión en paladar duro – VPH6

Uno de los resultados más significativos entre los obtenidos a lo largo de este trabajo, es el que se presenta en este caso. El mismo corresponde a un paciente con una lesión en paladar duro causada por VPH6. Dicha lesión había sido diagnosticada previamente tanto por test molecular (Captura de híbridos VPH-CH2), como por el estándar de oro (Biopsia convencional), y el interés estaba puesto en determinar la amplitud real del área afectada. Para ello se midieron los espectros de fluorescencia natural de la zona afectada y sus alrededores. En la figura 3.16 se muestran algunos de los círculos muestreados espectralmente en el paladar duro del paciente.

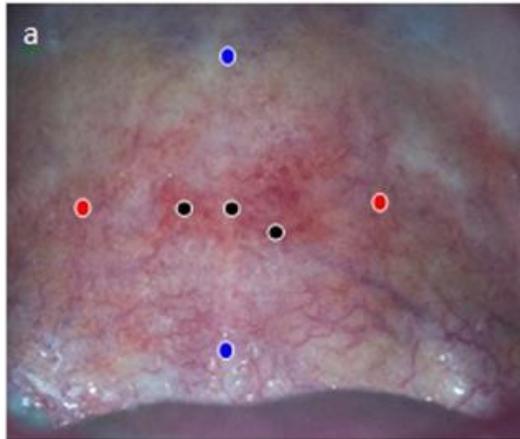


Figura 3.16 Fotografía del paladar duro, donde se indican con círculos coloreados los sitios de medición.

Los tres círculos marcados en negro fueron seleccionados dentro del área identificada por el odontólogo –Dr. Daniel Fabrizio– como clínicamente sospechosa, reconocida a simple vista por un marcado incremento de la intensidad de su tono rojizo respecto de su entorno. Los espectros de emisión de fluorescencia correspondientes se representan por trazos negros en la figura 3.17.

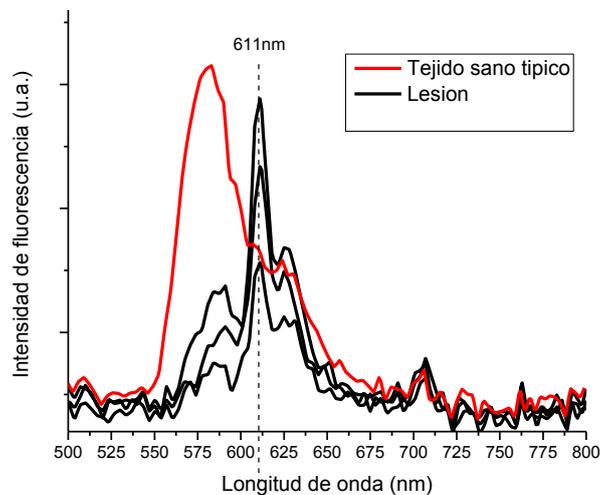


Figura 3.17 Espectros de fluorescencia registrados en las zonas indicadas con puntos negros en la figura 3.16.

En la figura 3.16, los cuatro círculos mostrados en color, dos en rojo y dos en azul, estarían ubicados en un área considerada sana a distancias de entre 0,5 y 1 cm del tejido afectado. Dichos espectros de fluorescencia se representan en la Figura 3.18 como trazos rojos y azules.

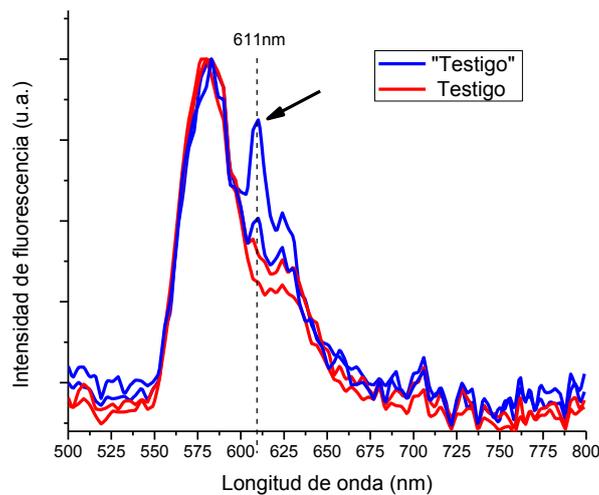


Figura 3.18 Espectros de fluorescencia registrados en las zonas indicadas con puntos rojos y azules en la figura 3.16.

Tal como puede observarse, los trazos azules poseen un pronunciado pico de emisión a 611 nm, tal como los que se representan en trazos negros de la figura 3.17. Este resultado indica que en el entorno de los círculos azules la infección por HPV6 está presente en forma intersticial por lo que su efecto macroscópico no puede observarse todavía visualmente. Este resultado observacional es un fuerte argumento para considerar que las biopsias ópticas por EEF deberían ser realizadas previamente a las biopsias convencionales, y fue oportunamente publicado por nuestro grupo de trabajo [56].

Como puede observarse en las Figuras 3.17 y 3.18, los espectros de fluorescencia de las regiones del paladar duro afectadas por el VPH6 presentan, además de otras características menos relevantes, el pico de 611 nm, mientras que las regiones sanas que la rodean tienen sus espectros típicos no lo incluyen. En este sentido, el resultado es similar al obtenido en el caso 5 presentado previamente, ya que en ese caso las señales provenientes de todas las zonas afectadas por los condilomas también presentan el característico pico de 611 nm.

Cabe destacar que si bien los espectros de fluorescencia natural registrados en las zonas de piel sana en torno a las lesiones del HPV6 perianal difieren notablemente de los correspondientes a la mucosa del paladar duro sano, para las zonas afectadas por VPH en ambos casos aparece el pico característico de 611 nm. En consecuencia este pico pareciera ser parte del espectro que identifica la infección por HPV6, o un indicio de su presencia. Esto se plantea aquí como un resultado preliminar, por lo que sería de interés poder abordar un estudio más exhaustivo a futuro, extendiendo además la cantidad de casos para lograr resultados concluyentes.

Para este paciente, como estudio complementario se realizó una inspección visual con ácido acético (IVA) al 3-5% y, tal como se muestra en la foto de la derecha de la figura 3.19, observó una reacción positiva luego de la topicación.



Figura 3.19 Fotografía del paladar duro del paciente antes (Izq.) y durante (Der.) la IVA.

Este procedimiento es habitual para la detección de lesiones en el cuello uterino. La IVA se realiza mediante un lavado del cuello uterino con una solución de entre 3% y 5% de ácido acético (vinagre). Luego de esperar aproximadamente un minuto, se observa directamente el cuello uterino, para identificar cambios de color en el mismo. Si se observan áreas blancas bien definidas (acetoblanco) cerca de la zona de transformación, se considera que la prueba es positiva para cambios celulares precancerosos o cáncer invasivo en su estadio temprano. En ese caso se recomienda realizar una segunda prueba para realizar el diagnóstico definitivo.

Es interesante destacar que, en este caso, la reacción acetoblanca positiva en el paladar duro que se observa en la figura 3.19 es una interesante prueba observacional de que la clásica biopsia óptica ginecológica podría ser extrapolada como biopsia óptica oral.

Caso 7. Epiteliomas basocelulares

El siguiente caso corresponde a una persona de 58 años de edad, que padecía cáncer de piel. Dicha persona generaba lesiones con frecuencia, que fueron tratadas por el Dr. Horacio Poteca a lo largo de 15 años mediante exéresis y fotodesnaturalización del lecho quirúrgico con laser de CO₂ en un principio, y luego con laser de diodo.

En el presente Trabajo de tesis se describe el tratamiento realizado sobre los epiteliomas basocelulares que presentaba en cabeza y en la cara mediante TFD. El FS utilizado en este caso fue ALA de aplicación local, por lo que sólo fue necesario proteger de la luz las zonas afectadas a diferencia de la aplicación sistémica. En la figura 3.20 se muestran los vendajes utilizados desde la aplicación de la droga hasta la iluminación, que se realizó 12 hs después.



Figura 3.20 Protecciones utilizadas en las áreas con FS aplicado.

Para delimitar el área a tratar y las zonas a proteger durante la TFD, se observó la fluorescencia inducida debida a presencia del FS administrado mediante la iluminación con un LED que emite a 400 nm. Asimismo se registraron EEm para cuantificar la concentración del FS en la lesión y en zonas circundantes. Este análisis se realizó tanto en la cabeza como en la cara del paciente y fue de suma importancia ya que a partir de estas observaciones se realizaron además los cálculos dosimétricos de la luz a administrar para activar el FS. La figura 3.21 muestra una fotografía de la cara del paciente durante el DFD. En ella puede observarse un área tumoral extensa e irregular en su nariz determinada por la observación de la fluorescencia inducida al ser iluminada por la luz LED de 400 nm.

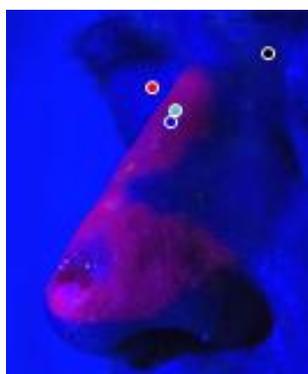


Figura 3.21 DFD: fotografía de la nariz del paciente, iluminada con un LED que emite en 400 nm, donde puede observarse el área tumoral determinada por la observación de la fluorescencia inducida.

El resultado de esta biopsia óptica generalizada de observación visual se verificó con los espectros de fluorescencia medidos sobre la piel del paciente en los puntos indicados por los círculos coloreados de la figura 3.21. Los espectros obtenidos se muestran en la figura 3.22, representados según los colores de la figura anterior.

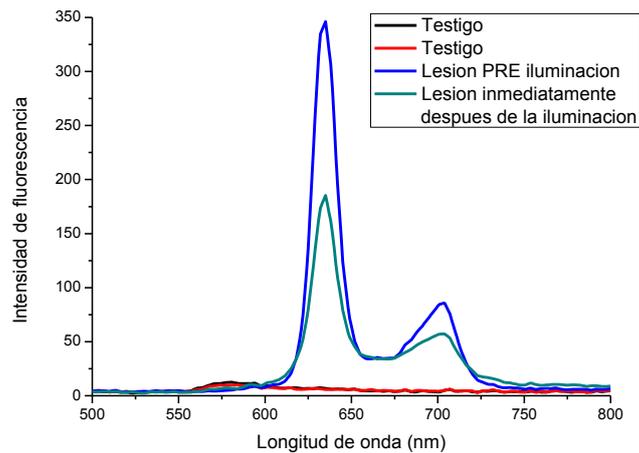


Figura 3.22 Espectros de emisión de fluorescencia registrados sobre la nariz del paciente del caso 7.

Los espectros registrados sobre la región tumoral antes y después de la TFD se muestran en azul y verde, respectivamente. Por otro lado, los espectros registrados sobre áreas sanas ubicadas a ambos lados de la nariz se muestran en negro y rojo. Nótese que los espectros de la región tumoral muestran una similitud muy marcada con el de la PpIX (figura 2.9). El hecho de que los registros tomados sobre la zona afectada sólo contengan los picos de 635 y 705 nm indica que la emisión fluorescente proviene de la PpIX.

En la figura 3.23 se muestran dos fotografías de la nariz del paciente luego de finalizado el tratamiento, y allí se pudo observar el enrojecimiento de la zona tratada. Es importante mencionar aquí que dicha zona obviamente se corresponde con el área que presenta fluorescencia roja en la figura 3.21 (DFD), debido a la absorción selectiva FS.

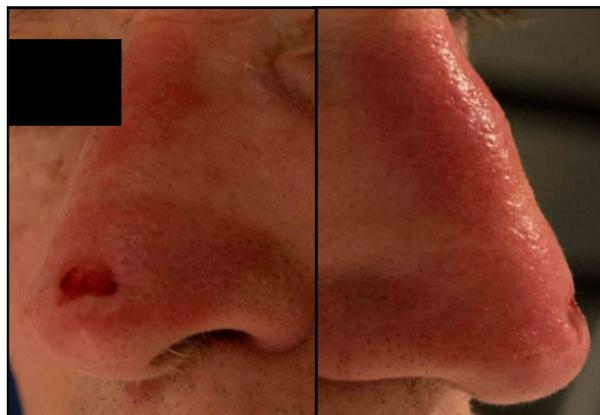


Figura 3.23 Fotografías del paciente luego del tratamiento.

Luego de la TFD se registraron nuevamente los espectros de emisión de fluorescencia en la región tratada, verificándose una significativa disminución de la emisión atribuida a la PpIX (curva verde en la figura 3.22).

Las lesiones en la cabeza de este paciente fueron tratadas con el mismo procedimiento. Sin embargo, en este caso no se usó el láser habitual que emite en 654 nm sino una lámpara de LEDs que emite en 637 nm con un ancho espectral de 18,1 nm (descrita en la sección 2.2). La figura 3.24 es una fotografía tomada durante el procedimiento, donde puede observarse la lámpara LED.



Figura 3.24 TFD en la cabeza del paciente, realizada con una fuente LED.

La figura 3.25 muestra dos fotografías de la cabeza del paciente, donde se ilustra el DFD realizado previo a la TFD. En la imagen de la derecha, puede observarse la zona a tratar delimitada selectivamente por la fluorescencia roja de la PpIX inducida por la iluminación con luz LED de 400 nm.

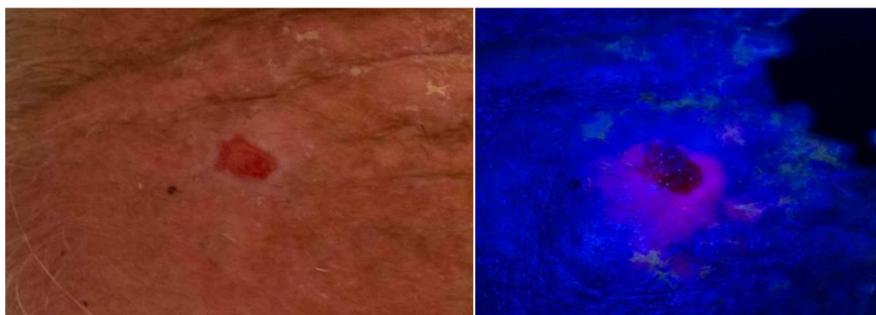


Figura 3.25 Diagnóstico fotodinámico en la cabeza del paciente.

Caso 8. Adenocarcinoma de parótida

El siguiente caso es un tratamiento de terapia fotodinámica realizado a un paciente de 52 años que padecía una recidiva de un adenocarcinoma de parótida con infiltración de parte de la hemilengua derecha.

Los tumores salivales pueden desarrollarse en cualquiera de las glándulas salivales (parótidas, submaxilares, sublinguales u otras de las entre 600 y 1000 glándulas salivales

menores). De ellos, el 80% se presentan en la parótida, el 10 a 15% en la glándula submaxilar y entre un 5 y 10% en la sublingual y glándulas salivales menores. Los adenocarcinomas son carcinomas epiteliales de alta malignidad que, cuando se presentan en la parótida, frecuentemente invaden al nervio facial. Producen metástasis ganglionares y, en el 50% de los casos, también a distancia. [57] La supervivencia a cinco años se acerca al 30%.

La figura 3.26 muestra una fotografía de la zona afectada junto con una imagen térmica de la misma, ambas tomadas con anterioridad al tratamiento. En la termografía IR se evidencia un notable incremento de la temperatura de la zona afectada respecto de la zona sana circundante. Esto se debe a la alta vascularización de las lesiones neoplásicas, y es otro ejemplo de la contribución de la termografía al diagnóstico.

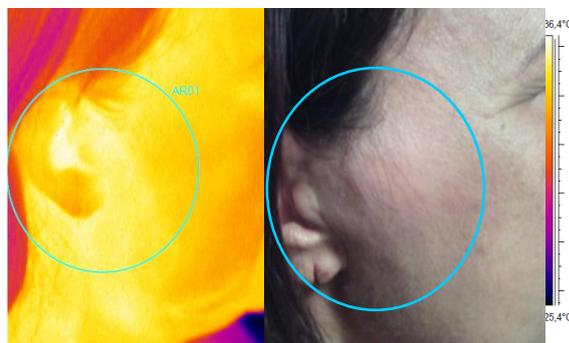


Figura 3.26: Termografía IR donde se observa la hipervascularización del carcinoma parotídeo en el hemifacial derecho del paciente, junto con una fotografía convencional.

En la figura 3.27 se observan los espectros de autofluorescencia (fluorescencia natural), de la zona afectada y sus alrededores. Allí también se utilizó la convención del reloj para describir la localización espacial de los sitios seleccionados para el análisis espectral.

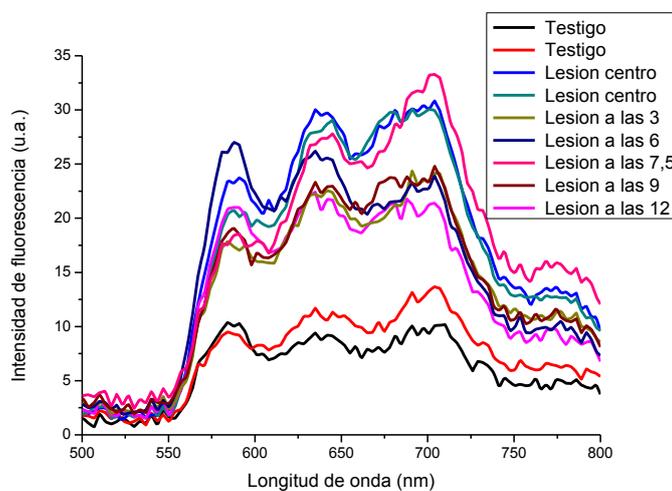


Figura 3.27 Espectros de fluorescencia natural para el Caso 8.

Los dos espectros medidos sobre zonas sanas, a modo de testigos, se representan con líneas negra y roja, y corresponden a puntos sobre el surco nasogeniano derecho y la articulación acromioclavicular derecha, respectivamente. En la figura se observa que estos dos espectros presentan una menor intensidad global de fluorescencia respecto de aquellos registrados en las zonas afectadas.

En la figura 3.28 se muestra la termografía IR realizada sobre el lado izquierdo del cuello, donde se observa el incremento de temperatura del ganglio superficial.

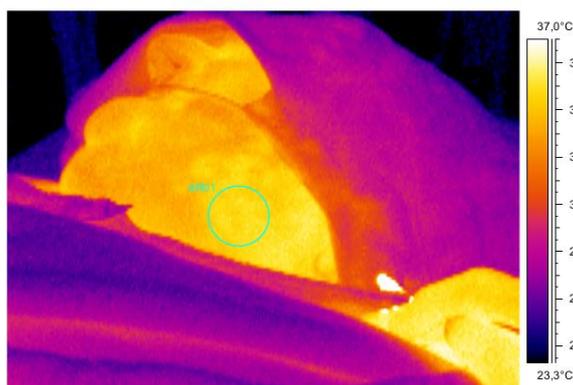


Figura 3.28 Termografía IR del ganglio izquierdo superficial.

Debido a la complejidad del caso, con compromiso del nervio facial y de las cuerdas vocales, un tratamiento quirúrgico podría haber resultado en una pérdida de movilidad y/o del habla. En ese contexto se decidió realizar el tratamiento con terapia fotodinámica. Cabe resaltar aquí que, debido a su selectividad, la TFD no produce lesiones en los nervios periféricos, por lo que se evita el riesgo de sufrir parálisis facial. En cambio, hay que tener especial cuidado en proteger los vasos de mayor calibre de la zona, ya que poseen restos de FS en circulación.

Una vez inyectada la droga fotoactivable (que en este caso fue el compuesto Foscan®) se registraron los espectros de fluorescencia inducida para corroborar la localización de la droga en el lugar y el área a tratar. En la figura 3.29 se muestran los espectros registrados en la zona de la boca. Obsérvese que la localización del FS (calculada a partir de la intensidad de fluorescencia registrada a 650 nm), es un 55% superior en la lesión respecto de la zona sana adyacente.

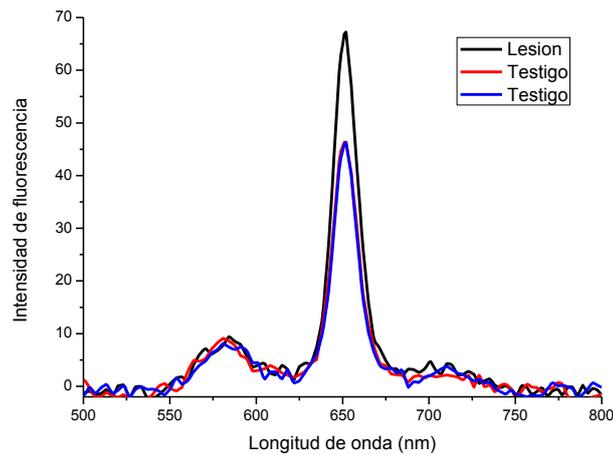


Figura 3.29 Espectros de fluorescencia inducida por la presencia del FS.

Posteriormente, se realizó la iluminación en quirófano con el láser de diodo (654 nm). Dada la gran extensión de la lesión, fue necesario iluminar desde varios lugares diferentes para activar el FS y cubrir la totalidad del área a tratar. En la figura 3.30 se observan fotografías tomadas durante el procedimiento. En la imagen de la izquierda se observa al médico interviniente iluminando a distancia, mientras que en la de la derecha se muestra cómo se realizó la iluminación por dentro de la cavidad bucal. En ambas puede observarse además que el procedimiento se realiza con muy poca luz ambiente, para evitar las complicaciones que pudieran surgir debido a la administración sistémica del FS.

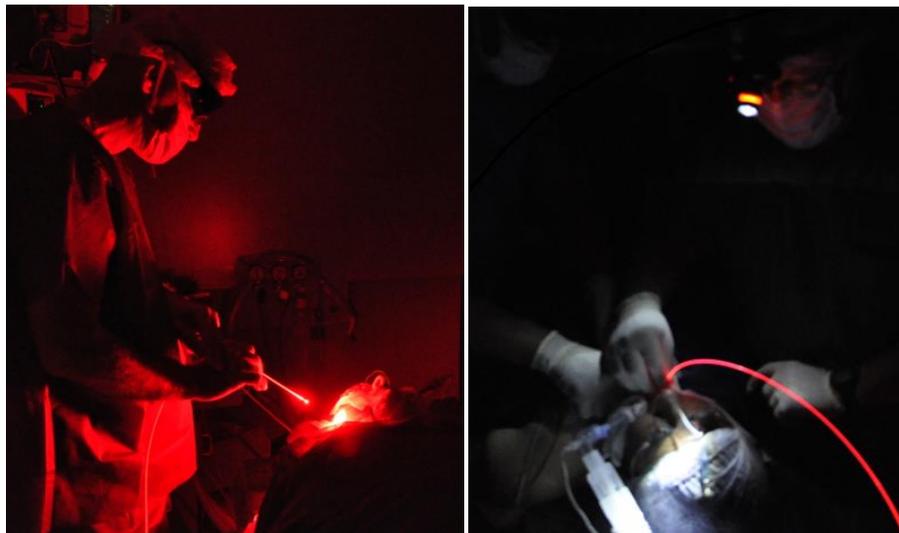


Figura 3.30 Iluminación con láser de diodo durante la TFD. (Izquierda) Iluminación externa. (Derecha) Iluminación por dentro de la cavidad bucal.

Se realizaron controles de área tratada minutos posteriores a la TFD, 24 horas después y luego de transcurridos 45 días. En las figuras 3.31 y 3.32 se muestran imágenes tomadas en el último control. En ellas puede verse la fístula remanente y el tejido en recuperación. La figura 3.31 muestra cómo quedó la zona tratada 45 días después del tratamiento y fue tomada mientras se registraban los espectros de fluorescencia de control.



Figura 3.31 Zona tratada, 45 días después de la TFD.

En la fotografía de la izquierda de la figura 3.32 se muestra la imagen de fluorescencia natural tomada sobre el vidrio de la lupa UV que se presentó en la sección 2.3. Esa imagen se realizó para verificar la completa eliminación del FS. La imagen de la derecha es una termografía IR de la misma zona, registrada con una placa termográfica de contacto, en la que sólo se observa un pequeño incremento de temperatura en los alrededores de la fístula. Es importante destacar aquí que en ésta termografía no se observa el gran incremento de temperatura dado por la hipervascularización tumoral, registrado en la realizada previamente al tratamiento (figura 3.26). Este resultado constituye otra evidencia del éxito del tratamiento.



Figura 3.32 Control realizado 45 días luego de la TFD. (Centro) Fotografía de la zona tratada. (Izq.) Imagen de fluorescencia natural tomada con la lupa UV. (Der.) Termografía IR de la misma zona.

Caso 9. Adenocarcinoma de colon – TFD por fibrocolonoscopia

El siguiente caso corresponde a una persona de 68 años de edad afectada por la presencia de pólipos en el colon, uno de los cuales fue tratado mediante TFD.

Los pólipos son prominencias de tejido que protruyen hacia la luz del órgano, es decir se extienden por sobre la mucosa que recubre interiormente el intestino grueso. Si bien la mayoría de ellos no son peligrosos, ciertos pólipos pueden convertirse en cancerosos e incluso ser malignos al momento de su detección. En el sentido por PET-CT (Tomografía por emisión de positrones y tomografía computada) realizado al paciente, uno de los pólipos del colon ascendente mostró una alta concentración de radionucleido. Para precisar el diagnóstico se le efectuó una biopsia convencional, que dio como resultado adenocarcinoma de colon. Entonces, se decidió tratarlo con terapia fotodinámica por vía colonoscópica.

El fotosensibilizador utilizado en esta oportunidad fue Foscan® y se administró por vía endovenosa. Pasadas las 96 hs que se requieren para su localización en el tumor se lo iluminó con el láser de diodo de 654 nm. En este caso la iluminación se realizó por vía endoscópica, para lo cual se utilizó una fibra óptica estéril de gran longitud que fue incorporada al canal de instrumental del colonoscopio. La figura 3.33 muestra dos fotografías tomadas antes (izquierda) y durante (derecha) el procedimiento. En la primera puede observarse el atuendo que viste el paciente, con todo su cuerpo cubierto para protegerse de la luz, y en la siguiente se observa al médico realizando el procedimiento, guiado por las imágenes endoscópicas.

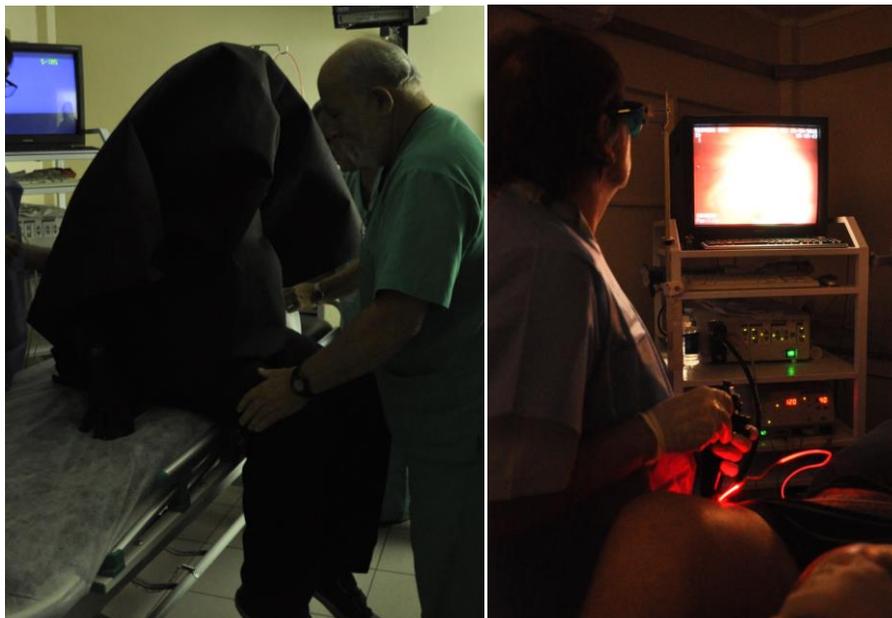


Figura 3.33 Fotografía del paciente: (Izquierda) momentos previos a la iluminación, donde se puede observar la vestimenta que utilizó para protegerse de la luz; (Derecha) durante

la iluminación por colonoscopia, donde puede observarse al médico guiando el procedimiento con las imágenes de la pantalla.

Cabe destacar que la iluminación del pólipo con la luz del diodo láser implicó una tarea compleja llevada a cabo por varios profesionales: el médico colonoscopista, que se encargó de mantener el extremo del colonoscopio apuntado al pólipo permanentemente compensando los movimientos peristálticos naturales y aleatorios del colon; el Dr. Poteca, quién se ocupó de mantener el extremo de la fibra óptica que conduce la luz del láser a una distancia constante del pólipo, indicando, además al operador del láser variar la intensidad de emisión de acuerdo a la circunstancia, y por último el operador del láser.

La figura 3.34 muestra una imagen endoscópica tomada durante la TFD, donde se observa la punta de la fibra óptica y la luz roja proveniente del laser de diodo.

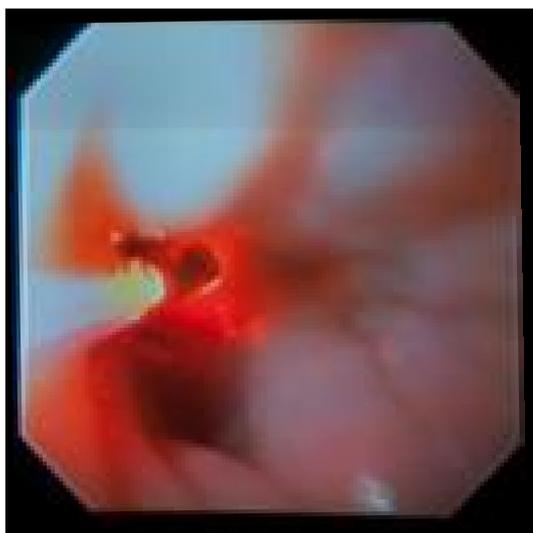


Figura 3.34 Imagen endoscópica tomada durante el procedimiento.

En este caso no fue posible medir los espectros de emisión de fluorescencia sobre el pólipo debido a la imposibilidad de acceder al sitio con la fibra óptica del fluorómetro utilizado en el presente trabajo. Sin embargo, sesenta días después de realizada la TFD, se registraron espectros de emisión de fluorescencia sobre la piel del paciente a los fines de sensor la cantidad de fotosensibilizador remanente su cuerpo. Para ello se tomó como punto representativo del lugar de mayor concentración de droga, aquél donde se realizó la aplicación de la inyección. Las curvas registradas sobre el pliegue del codo (lugar donde se había inyectado la droga), y en zonas cercanas se muestran en la figura 3.35.

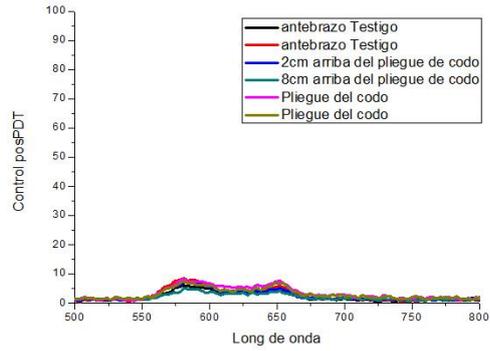
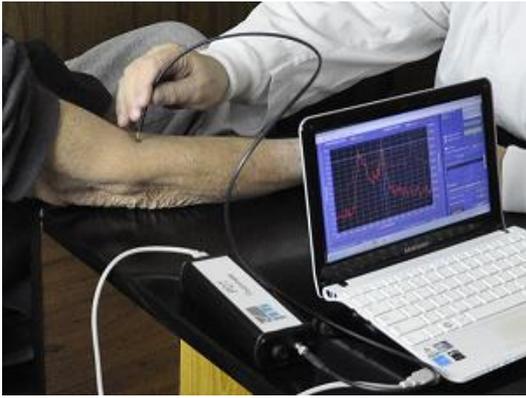


Figura 3.35 Fotografía que muestra el registro de los espectros de emisión (a la izquierda), y las curvas obtenidas (derecha).

En ella se observa que no existe ninguna diferencia apreciable entre las curvas correspondientes al pliegue del codo y los restantes, además de la ausencia de los picos característicos del FS. De esta manera se verificó la ausencia de fotosensibilizador 60 días luego de la TFD.

Parte II

Evaluación de la utilización de la autofluorescencia de muestras de orina en la detección de enfermedades neoplásicas

Capítulo 4. Materiales y métodos empleados para el estudio de fluorescencia de orinas

4.1 Recolección y tratamiento de muestras previo a la obtención de información espectroscópica

4.1.1 Obtención de las muestras

Las muestras de orina incluidas en el presente trabajo de Tesis fueron recibidas de personas mayores de edad en todos los casos. Con fines comparativos se consideraron pacientes con diagnóstico de patologías oncológicas, personas sanas y pacientes con otras patologías.

Todas las personas participaron de manera voluntaria, se les pidió que dieran su consentimiento por escrito y completaran una ficha médica resumiendo las condiciones en las que se encontraban al momento de colectar la muestra. Los datos solicitados en la ficha incluyeron género, edad, diagnóstico oncológico, y preguntas tales como: ¿Es fumador?, ¿Ingiere algún tipo de medicación? (¿Cuál?), ¿Posee alguna de las siguientes patologías: diabetes, infección bacteriana, ictericia, hipertensión arterial, alguna infección oportunista activa? ¿Posee alguna otra co-morbilidad?, ¿Está embarazada o en período de lactancia? Cantidad de horas de ayuno y cantidad de horas de retención al momento de coleccion. Todos estos datos se registraron para cada una de las muestras, y a partir de esa información se las agrupó convenientemente para su análisis. Se trabajó en el protocolo según las normas establecidas de Protección de Datos Personales (Ley 25.326). En ningún caso se solicitaron datos personales ni identificatorios de los pacientes.

En una primera etapa de análisis, se requirió que las muestras a incluir cumplieran ciertas condiciones. Esas condiciones iniciales se listan a continuación:

- Que la orina a colectar fuera la primera de la mañana, con un tiempo de retención mayor a 5hs.
- Estar en ayunas al momento de colectar la muestra y no haber ingerido ninguna medicación durante las 24 horas previas.
- Para participar del estudio, las personas afectadas con patologías oncológicas debían tener el diagnóstico confirmado, no estar bajo ningún tratamiento y no podían estar cursando enfermedades coexistentes o adicionales en relación con el diagnóstico oncológico (co-morbilidades).
- Las personas no afectadas por patologías oncológicas tampoco debían estarlo por ninguna otra enfermedad.

Luego, debido principalmente a la dificultad en la obtención de muestras, se amplió el estudio incluyendo muestras que no cumplieran algunos de estos requisitos. En el Capítulo 5 del presente estudio, se describe el análisis matemático de los datos registrados y se precisan los criterios de inclusión de las muestras en cada caso.

Para la medición tanto de los espectros de fluorescencia como de las matrices de excitación/emisión, se utilizaron entre 3 y 5 ml del sobrenadante de las muestras tomadas para realizar el análisis de “orina completo”, solicitado para control de rutina o para evaluación del curso de alguna patología. Todas las muestras, que se recolectaron en recipientes estériles, fueron conservadas en la heladera y al reparo de la luz directa hasta su utilización dentro del mismo día de colección. Esta ventana temporal de 24hs post colección se fijó luego verificar la constancia de las señales espectroscópicas de cada muestra en ese lapso de tiempo. Para ello se analizó la fluorescencia de la orina transcurridos diferentes tiempos desde la colección (entre 20 minutos y 24 horas), habiendo obtenido resultados equivalentes.

4.1.2 Pre-tratamiento de las muestras

Reactivos

Para diluir las muestras, en los casos que fue necesario, se utilizó un buffer fosfato de $\text{pH} = 6,2$ ($\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$ 0,001 M). Dicho buffer se preparó disolviendo las cantidades apropiadas de Na_2HPO_4 y de NaH_2PO_4 (Biopack® PA) en agua bidestilada. Para evitar la formación de microorganismos el buffer fue filtrado con membrana de acetato de celulosa de 0,22 micrómetros de poro y conservado hasta el momento de su uso a 4° durante no más de 15 días. Además de tomar estos recaudos, periódicamente se realizaron MEE al buffer, de modo de verificar la conservación de las condiciones óptimas (sin descomposición apreciable y libre de impurezas que puedan afectar la medida).

En los casos que fue necesario rectificar el pH de las muestras que no necesitaban dilución, se utilizó una solución de ácido acético 0,1 M, para llevar el pH a valores en el rango comprendido entre 5,5 y 6,5.

Para el análisis de las matrices de excitación-emisión, que se presenta en el Capítulo 6, se utilizó como blanco la matriz del agua bidestilada (ver sección 4.3.4).

Centrifugación

Si bien en los casos en los que se utilizó el sobrenadante de las muestras para el análisis convencional de orina completa la muestra ya había sido centrifugada, todas las muestras analizadas se centrifugaron durante 10 minutos a 3000 RPM inmediatamente antes de registrar sus espectros. Esto se realizó con el fin de eliminar cualquier elemento dispersor de luz que pudiera haberse generado por sedimentación durante su

almacenamiento. Este procedimiento se realizó con todas las muestras, independientemente de si habían sufrido una centrifugación previa o no.

Dilución

La concentración de metabolitos en la orina es sumamente variable ya que depende de: el tiempo de retención, la cantidad de líquido ingerido por el paciente, los alimentos que ingiere, el estado de salud del paciente y varios factores más que son muy difíciles de cuantificar y estandarizar para todo el procedimiento. En este trabajo no se intentó medir la concentración, pero es necesario tener en cuenta los efectos que la misma puede tener sobre la absorción total de la radiación por parte de la muestra para evitar los efectos de filtro interno en las medidas de fluorescencia. Los efectos de filtro interno ocurren cuando parte de la radiación incidente y/o la emitida es absorbida por las especies presentes en la muestra, de manera que se reduce la cantidad de radiación efectiva que llega al detector. Estos efectos influyen notablemente en la forma de los espectros de emisión de muestras que contienen cromóforos en alta concentración, como es el caso de la orina, y existen varios métodos para corregirlos. En el presente estudio han sido corregidos mediante el empleo de los correspondientes espectros de absorción usando la siguiente ecuación:

$$I_{Corr} = I_{Reg} \times 10^{\frac{Abs_{Em} + Abs_{Ex}}{2}}$$

donde I_{Corr} es la intensidad de emisión corregida, I_{Reg} es la intensidad de emisión registrada, Abs_{Ex} corresponde la absorbancia de la muestra a la longitud de onda de excitación (causa del efecto de filtro interno primario) y Abs_{Em} corresponde la absorbancia de la muestra a la longitud de onda de emisión (causa del efecto de filtro interno secundario, también denominado autoapagamiento). En los casos en los que la absorbancia de la muestra fue mayor a 1,8 para alguna longitud de onda dentro de rango de interés, la muestra se diluyó hasta obtener valores menores con el fin de asegurar un comportamiento lineal. Esta diluciones se realizaron con el buffer de fosfato de pH = 6,2 descrito previamente en esta sección.

Para los espectros de emisión de fluorescencia EEm que se registraron excitando con luz de 405nm, en la mayoría de los casos no fue necesario diluir las muestras. Además, en los pocos casos en los que se requirió, ninguna dilución superó 1 parte de orina por cada 3 partes de buffer. En cambio, éste fue un paso fundamental para trabajar en el rango de linealidad cuando se registraron las matrices de emisión-excitación (MEE), teniendo que diluir entre 150 y 700 veces dependiendo de la muestra en particular.

4.1.3 Pretratamiento de los datos

Normalización.

La normalización de un espectro de fluorescencia implica multiplicar (y/o dividir) un espectro completo por algún factor. Si bien este procedimiento elimina la información relacionada con la intensidad absoluta, trabajar con espectros normalizados posee la ventaja de que no es necesario registrar la intensidad de fluorescencia en unidades calibradas. Además en muchos casos facilita la comparación entre espectros y no afecta a las medidas relativas. En el presente trabajo, según el caso, se normalizaron los datos de tres maneras diferentes:

Normalización al valor máximo: cada espectro de fluorescencia se normalizó dividiendo el valor observado de la intensidad de fluorescencia en cada longitud de onda de emisión, por el valor de la máxima intensidad de emisión de ese espectro. Luego cada valor obtenido se multiplicó por 1000. De esta manera todos los espectros obtenidos tienen el mismo valor de intensidad máxima de fluorescencia (igual a 1000 unidades arbitrarias).

Normalización a la misma área: cada espectro de fluorescencia se normalizó dividiendo el valor observado de la intensidad de fluorescencia en cada longitud de onda de emisión por la suma total de las intensidades de emisión de fluorescencia registradas en todo el rango espectral analizado. Dada la resolución espectral empleada, dicho valor se corresponde prácticamente con el área que se encierra bajo la curva del espectro. De esta manera los espectros normalizados tienen el mismo valor de intensidad total de fluorescencia (igual a 1).

Normalización Euclídea: cada espectro de fluorescencia se normalizó de manera tal que la raíz cuadrada de la sumatoria de todos sus valores elevados al cuadrado sea igual a uno. Esto es equivalente a pensar cada espectro como un vector n-dimensional (con n igual al número de longitudes de onda), de módulo unitario. La “distancia” euclídea (caso general de N dimensiones) en general es un parámetro más robusto que la sumatoria convencional de cada elemento, ya que la suma está compuesta sólo de valores positivos (por estar elevados al cuadrado).

4.2 Técnicas analíticas

4.2.1 Espectrofotometría UV-Vis

Esta técnica analítica está basada en la medida directa de la absorción de radiación electromagnética por parte de una muestra. En este trabajo los espectros de absorción UV-visible se obtuvieron con el espectrofotómetro UV-Vis Agilent Cary 60 que se muestra en la figura 4.1. Dicho equipo posee programas adecuados para registrar y almacenar los

espectros, una lámpara Flash de Xenon con un rango espectral es 190 – 1100 nm y, de acuerdo con el fabricante, el equipo tiene un intervalo fotométrico entre -4 y 4 unidades de absorbancia. La resolución espectral empleada fue 1,5 nm y se obtuvieron espectros de manera automática fijando previamente el intervalo de longitudes de onda entre 200 y 900 nm.

Los espectros se registraron utilizando agua bidestilada como blanco y se emplearon celdas de cuarzo de 10 mm de camino óptico. Tal como se mencionó previamente, las soluciones cuyas absorbancias en el rango de interés presentaron valores mayores que 1,8 fueron diluidas para evitar pérdida de linealidad en las determinaciones espectroscópicas.



Figura 4.1 Espectrofotómetro UV-Vis utilizado.

4.2.2 Espectrofluorimetría UV-Vis

Los espectros de emisión de fluorescencia (EEm) y las matrices de excitación y emisión (MEE) de las muestras de orina se obtuvieron usando los espectrofluorómetros Varian Cary Eclipse (Fig 4.2) y Perkin Elmer LS50B. Ambos instrumentos tienen la capacidad de operar en el rango de excitación/emisión de 200 a 800 nm.

Para obtener los EEm las muestras fueron excitadas con radiación de 405 nm y se barrió el rango de longitudes de onda de emisión comprendido entre 425 y 700 nm a intervalos de 2 nm. La velocidad de escaneo fue fijada en 240 nm/min y el tiempo de integración en 0,5 s.

Para el registro automático de las matrices de excitación/emisión (MEE) se utilizó el espectrofluorómetro Varian Cary Eclipse. En este caso, se barrió el rango de longitudes de onda de excitación de 250 a 450 nm, a intervalos de 5 nm, y de 275 a 600 nm para las longitudes de onda de emisión, a intervalos de 2 nm. La velocidad de escaneo fue fijada en

240 nm/min y el tiempo de integración fue 0,5 s. Los datos correspondientes a cada muestra se almacenaron en matrices de excitación-emisión.

Para las medidas de fluorescencia, las muestras con valores de absorbancia mayores a 1,8 también fueron diluidas. Las medidas de emisión se realizaron con celdas de cuarzo de 1 cm de camino óptico, con las cuatro caras transparentes. Ambas rendijas (de emisión y de excitación) se mantuvieron constantes a lo largo de todo el trabajo en 5 nm.

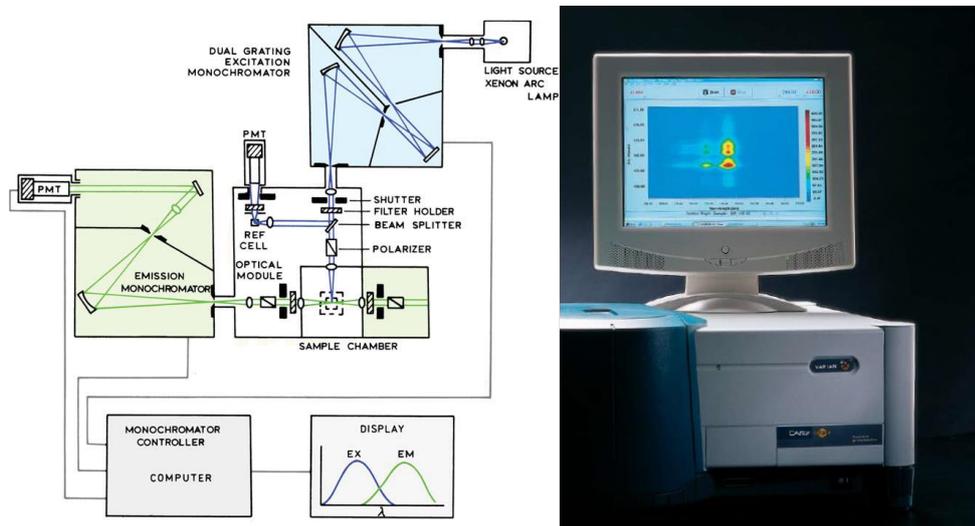


Figura 4.2 Espectrofluorómetro utilizado.

4.3 Análisis estadístico y test diagnósticos

Si bien cada espectro de emisión de fluorescencia contiene mucha información, gran parte de ella podría no ser relevante para el objetivo de este estudio. En este contexto, en los Capítulos 5 y 6 del presente trabajo de Tesis, se presenta la búsqueda y el análisis de diferentes valores representativos de los espectros (que llamaremos índices de interés), para comparar y tratar de establecer si existen diferencias en la fluorescencia de las muestras de orina entre personas sanas y de personas afectadas por patologías oncológicas. En realidad, en un primer paso se intentó determinar si algún índice obtenido a partir de la fluorescencia de la orina era capaz de reflejar la existencia de dos poblaciones, para luego, en un paso posterior determinar cuál, de entre aquellos índices de interés, presenta la mejor capacidad de clasificación. Por lo tanto, para realizar las comparaciones numéricas entre poblaciones, resultó fundamental realizar un correcto y exhaustivo análisis estadístico de los datos espectrofluorométricos.

En un estudio estadístico la elección de un método de análisis apropiado para los datos obtenidos depende de múltiples factores. En primer lugar depende de la forma en la los datos hayan sido obtenidos. Esto se refiere a si las observaciones corresponden a dos grupos independientes de individuos, o por el contrario si cada serie de datos se recoge en

los mismos sujetos bajo condiciones diferentes. En particular, para el caso del presente trabajo, el objetivo central fue comparar los valores de una variable continua (índice) registrados para dos poblaciones (sanos y enfermos). Las muestras de orina y todos los parámetros que se extrajeron de la información de sus espectros corresponden a personas individuales y esos datos se compararon entre diferentes personas, por lo que resulta obvio que las observaciones corresponden a grupos independientes.

Otro aspecto que a considerar es el tipo y la distribución de los datos. Para grupos independientes, los métodos paramétricos requieren que las observaciones en cada grupo provengan de una distribución aproximadamente normal y con una variabilidad semejante. Por lo tanto, al momento de seleccionar la metodología de comparación es necesario analizar si los datos verifican esas condiciones. La comprobación de la hipótesis de normalidad de las observaciones puede realizarse por métodos gráficos (histogramas, por ejemplo) o mediante tests estadísticos (test de Kolmogorov-Smirnov, test de Shapiro-Wilks). De no cumplirse esta hipótesis, se puede recurrir a una transformación de los datos (aplicación del logaritmo, raíz cuadrada, etc.), o a la utilización de procedimientos no paramétricos. De todas maneras, más allá de las comprobaciones, se considera que en el caso de poseer un número suficiente de observaciones, (mayor de 30), queda justificada la utilización de un método paramétrico bajo la hipótesis de normalidad.

Cabe destacar que, para todos los estudios sobre muestras de orina presentados en esta Tesis, el número de observaciones fue superior a 30 por lo que de lo expuesto anteriormente se podría asumir como válida la hipótesis de normalidad. No obstante ello, tal como se muestra en el Capítulo 5, esta suposición fue verificada explícitamente utilizando histogramas.

Una vez comprobada la hipótesis de normalidad es necesario analizar si la varianza es similar en ambos grupos de observaciones, es decir la homogeneidad de las varianzas (homocedasticidad). El método estadístico utilizado en el presente trabajo para decidir si las varianzas en ambos grupos son iguales o no fue el “test F” o razón de varianzas. Según dicho método, bajo la suposición de que las dos poblaciones siguen una distribución normal y tienen igual varianza se espera que la razón de las varianzas muestrales F (ecuación 4.1) siga una distribución F de Snedecor (o Fisher) con parámetros (n-1) y (m-1).

$$F = \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}{\frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2} = \frac{\hat{s}_x^2}{\hat{s}_y^2} \quad (4.1)$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (4.2)$$

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i \quad (4.3)$$

donde X_i representa el i-ésimo valor observado de la muestra de la población X (sanos por ejemplo), \bar{X} el valor medio de los valores observados y n el número total de elementos del grupo X (muestras). De manera similar Y_i , \bar{Y} y m representan el i-ésimo valor observado, el valor medio de los valores observados y el número total de elementos de muestra de la población Y (enfermos), respectivamente.

El valor de la razón de varianzas, o valor F obtenido debe compararse con el valor crítico de la distribución $F_{(n-1)(m-1)}^\alpha$. Si el valor del estadístico F calculado es mayor que el F crítico se rechaza la hipótesis nula, que asume homocedasticidad, con un nivel α , y se adopta la hipótesis alternativa que postula que las varianzas no son iguales (no existe homocedasticidad).

En función del resultado anterior se decide qué test estadístico utilizar para comparar las medias y determinar si su diferencia es estadísticamente significativa:

- Si las varianzas son iguales se utilizará el test de Student (o test T).
- Si las varianzas no son iguales se utiliza la prueba t de Welch.

En ambos casos, como hipótesis de partida (hipótesis nula, o H_0), se asume que el índice que se está analizando tiene el mismo valor para las dos poblaciones, es decir que las medias de dicho índice para personas sanas y enfermas toman los mismos valores. Posteriormente se evalúa la probabilidad de haber obtenido los datos observados si esa hipótesis es correcta. El valor de esta probabilidad coincide con el valor-p que proporciona cada test estadístico. Cuanto menor sea p, menos probable será que siendo cierta la hipótesis nula, se obtenga ese resultado por efecto del azar.

Test de Student:

En el caso en que se cumpla la hipótesis de homocedasticidad se emplea este test, que para dos muestras independientes se basa en el siguiente estadístico:

$$t_{st} = \frac{(\bar{X} - \bar{Y})}{\sqrt{\hat{S}^2 \left(\frac{1}{n} + \frac{1}{m} \right)}} \quad (4.4)$$

con \hat{S}^2 la varianza combinada definida como:

$$\hat{S}^2 = \frac{(n-1)\hat{S}_x^2 + (m-1)\hat{S}_y^2}{n+m-2} \quad (4.5)$$

Si la hipótesis nula (H_0) es cierta, el estadístico t_{st} sigue una distribución t de Student con $n+m-2$ grados de libertad. En ese caso, el valor obtenido debería estar dentro del rango de mayor probabilidad según esta distribución. Por el contrario, si el valor obtenido del estadístico es mayor al valor crítico correspondiente (según tabla para $n+m-2$

grados de libertad y un valor α determinado), se concluye que existen diferencias significativas entre las medias muestrales, y por lo tanto se rechaza la Hipótesis nula.

Test t de Welch:

En el caso en el que los dos grupos de observaciones independientes posean diferentes varianzas, se utiliza una modificación del test t de Student descrito anteriormente. Esta prueba es conocida como prueba de Welch y está basada en el estadístico t_w :

$$t_w = \frac{(\bar{X} - \bar{Y})}{\sqrt{\left(\frac{\hat{s}_x^2}{n} + \frac{\hat{s}_y^2}{m}\right)}} \quad (4.6)$$

Que bajo la hipótesis nula (H_0), sigue una distribución t de Student con f' grados de libertad, dados por:

$$f' = \frac{\left(\frac{\hat{s}_x^2}{n} + \frac{\hat{s}_y^2}{m}\right)^2}{\frac{1}{n-1}\left(\frac{\hat{s}_x^2}{n}\right)^2 + \frac{1}{m-1}\left(\frac{\hat{s}_y^2}{m}\right)^2} - 2 \quad (4.7)$$

El análisis se realiza de manera similar al test de Student: si la hipótesis nula es cierta, el estadístico t_w sigue una distribución t de Student con f' grados de libertad. La principal diferencia radica en que los grados de libertad para este caso dependen de los valores de las varianzas muestrales. Si el valor obtenido de estadístico es menor al valor crítico de tabla (según tabla para f' grados de libertad y α determinado), se acepta la hipótesis nula. Por el contrario, si el valor obtenido del estadístico es mayor al valor crítico correspondiente se concluye que existen diferencias significativas de las medias muestrales (se rechaza la Hipótesis nula).

Test diagnósticos:

El diagnóstico médico es un proceso dinámico en el que se intentan tomar las decisiones más apropiadas estando siempre en presencia de un cierto nivel de incertidumbre. Desde un punto de vista funcional, se considera prueba diagnóstica a cualquier procedimiento realizado para confirmar o descartar un diagnóstico o incrementar o disminuir su verosimilitud⁵. La utilidad de una prueba diagnóstica depende fundamentalmente de su efectividad y de su confiabilidad, pero también depende de otros factores que escapan a los fines del presente trabajo, como son su rendimiento clínico y su costo. En este sentido es importante mencionar que si bien en el presente trabajo no se realiza un análisis de costos, la espectroscopia de fluorescencia se plantea como un método sencillo y económico que permitiría desarrollar un test muy accesible.

⁵ El concepto "Verosimilitud" se refiere a la cercanía de una teoría a la verdad. Mide el mayor o menor contenido de verdad de las teorías o enunciados.

El empleo de tests estadísticos para predecir la presencia o ausencia de enfermedades es muy importante en investigación médica. Toda investigación sobre pruebas diagnósticas parte de la certeza de que una enfermedad existe o no en un grupo de individuos de una muestra. Es necesario entonces contar con una prueba diagnóstica suficientemente acreditada, que identifique el diagnóstico verdadero. Este patrón de referencia, conocido como "estándar de oro", se definió en el Capítulo 1 y puede ser una sola prueba o bien una serie de pruebas. Los resultados de la prueba diagnóstica en evaluación ("test") se comparan con los resultados de dicho patrón y se calculan ciertos parámetros para proporcionar una comparación objetiva de dos o más técnicas.

Los test diagnósticos que tienen sólo dos categorías posibles de resultados (positivo y negativo, por ejemplo), se denominan dicotómicos. En los casos en que tanto el estándar de oro como el test presentan resultados dicotómicos su enfrentamiento se puede resumir en una tabla de contingencia de 2 x 2 (2 filas y 2 columnas) donde la prueba patrón suele colocarse en columnas y el test a evaluar en filas (figura 4.3).

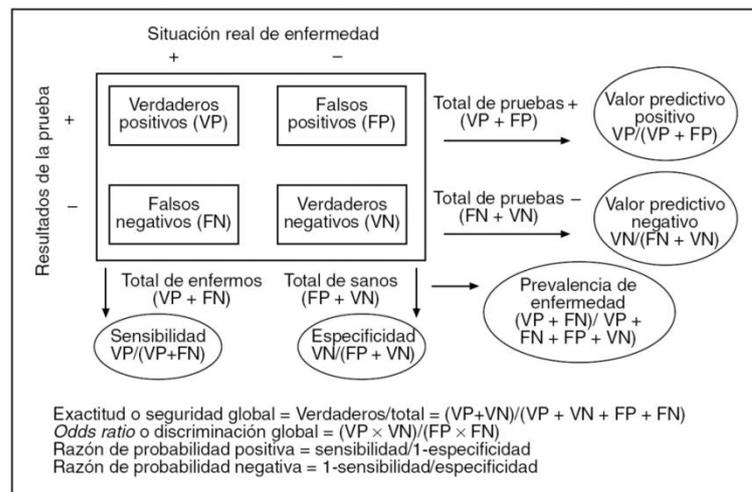


Figura 4.3 Tabla de contingencia para la evaluación de una prueba diagnóstica de resultado dicotómico. Adaptada de [58]

A lo largo del presente estudio todas las pruebas diagnósticas que se analizaron fueron dicotómicas, ya que tanto los resultados del estándar de oro como los de los tests sólo aceptaban dos clasificaciones posibles. Es decir, se clasificaron como "sanos" a aquellos que dieron resultado negativo a la prueba diagnóstica y como "enfermos" a los que dieron resultado positivo.

Entonces, a partir de la comparación de los resultados obtenidos del estándar de oro y del test en estudio se definen las siguientes categorías que fueron presentadas previamente en la tabla de contingencia:

- **Verdadero positivo (VP):** representa la clasificación de ambas técnicas como positivo (o enfermo).
- **Falso positivo (FP):** corresponde a una clasificación como positivo (enfermo) por la técnica en estudio (espectroscopia de fluorescencia, en este caso), y negativo (sano) por el estándar de oro (histología).
- **Falso negativo (FN):** corresponde a una clasificación negativa (sano) por la técnica en estudio (fluorescencia), y una clasificación positiva (enfermo) por el estándar.
- **Verdadero negativo (VN):** corresponde a la clasificación por ambas técnicas como negativo (sano).

Una vez que se han realizado las clasificaciones apropiadas, a partir de estas cantidades se calculan los indicadores de validez del test: *la sensibilidad*, que es la probabilidad de que la prueba resulte positiva si la enfermedad o condición de estudio está presente; *la especificidad*, que es la probabilidad de que la prueba resulte negativa si la enfermedad está ausente; los *valores predictivos positivos y negativos*, que corresponden la probabilidad de tener la enfermedad si la prueba ha resultado positiva y de no tener la enfermedad si la prueba ha resultado negativa, respectivamente; y la *precisión* global del test, que es la fracción de casos en los que test clasifica correctamente a un paciente.

La sensibilidad y la especificidad son características intrínsecas de la prueba diagnóstica que brindan información acerca de su validez antes de conocer el resultado. En cambio, los valores predictivos son probabilidades del resultado, es decir, dan la probabilidad de padecer o no la enfermedad una vez conocido el resultado de la prueba diagnóstica. Son valores post-test y además dependen de la prevalencia de la enfermedad⁶, es decir del porcentaje de una población que está afectada por esa determinada patología.

En particular, en este trabajo nos centramos en tres de los indicadores previamente mencionados: la sensibilidad, especificidad y precisión, que se calculan de la siguiente manera:

$$S = \textit{Sensibilidad} = \frac{VP}{VP+FN} \quad (4.8)$$

$$E = \textit{Especificidad} = \frac{VN}{VN+FP} \quad (4.9)$$

⁶ Prevalencia de una enfermedad se refiere a la tasa de incidencia de dicha enfermedad sobre la población global.

$$P = \textit{Precisión} = \frac{VP+VN}{VP+VN+FP+FN} \quad (4.10)$$

Es importante tener en cuenta que, en todos los casos, la información de la que se dispone sobre la validez de las pruebas diagnósticas procede de estudios realizados en muestras de una población y no sobre la población total. Es por ello que las estimaciones obtenidas en dichos estudios están sujetas a una cierta variabilidad aleatoria (por lo que resulta necesaria la aplicación de intervalos de confianza) e incluso a sesgos si los estudios no han sido diseñados correctamente.

Los estimadores de validez mencionados pueden aplicarse a pruebas con resultados discretos con más de dos categorías e incluso a resultados asociados a variables continuas. En el caso de tener variables continuas se pueden establecer puntos de corte para convertir el resultado a una variable discreta con dos o más categorías.

Con el objeto de explorar la capacidad diagnóstica de una prueba y a la vez analizar los distintos puntos de corte, frecuentemente se emplean las denominadas curvas ROC (iniciales del término inglés original: Receiver Operating Characteristics), ya que en forma bastante directa permiten realizar un análisis costo/beneficio en la toma de decisiones diagnósticas.

Una curva ROC es un gráfico que presenta en una forma visual sencilla cómo la sensibilidad y la especificidad de un test varían a medida que se modifica el valor umbral o de corte definido para la clasificación de las poblaciones. Es decir, si bien la sensibilidad y la especificidad son dos valores que miden la validez inherente a un test diagnóstico basado en resultados dicotómicos, cuando la prueba que se evalúa se mide en una escala cuantitativa ambos indicadores de validez varían en función del valor de corte (umbral) seleccionado para separar las poblaciones. Al variar el umbral, la sensibilidad crece a medida que la especificidad decrece, y viceversa.

El “valor óptimo” para el umbral se define como aquél que brinda la máxima clasificación correcta. En general, el valor umbral se selecciona de modo tal de darle el mismo peso a la sensibilidad que a la especificidad. Sin embargo, según el propósito del test que se está evaluando, puede ser de interés seleccionar el punto de corte desplazado de su punto medio con la intención de maximizar sensibilidad a expensas de reducir especificidad, o viceversa. Por ejemplo si se busca disponer de un test diagnóstico para “screening” de una población, como primer test de sospecha de enfermedad, se buscará que el mismo sea altamente sensible de manera que ningún enfermo sea clasificado como sano. En este caso se deberá tener en cuenta que el resultado “positivo” indicará la necesidad de un segundo test que confirme la enfermedad. Por el contrario, si el test se requiere para la confirmación de una patología se buscará que sea altamente específico, y

no tan sensible. Hay que tener en cuenta que en cualquiera de estos dos casos se puede estar limitando la aplicabilidad de la prueba ya que dichos valores sólo le serán útiles a una parte de los pacientes que se sometan a ella.

Tal como se mencionó anteriormente, la curva ROC es un gráfico bidimensional en el que se observan todos los pares sensibilidad/especificidad resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados (Figura 4.4).

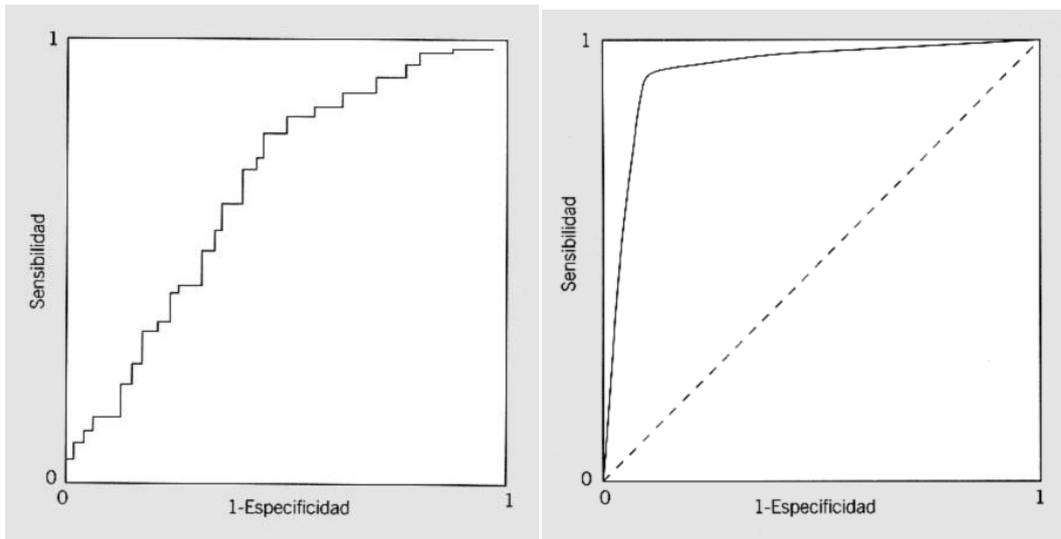


Figura 4.4: Curvas ROC representativas. Adaptado de [59]

El eje X representa la fracción de falsos positivos, o (1-especificidad), y en el eje Y se representa la fracción de verdaderos positivos (sensibilidad). Cada punto de la curva representa un par $(1-E, S)$ correspondiente a un punto de corte determinado. Una prueba con discriminación perfecta, sin solapamiento de resultados en las dos poblaciones, tiene una curva ROC que pasa por la esquina superior izquierda, donde S y E toman valores máximos (S y $E = 1$). Una prueba sin discriminación, con igual distribución de resultados en los dos subgrupos, da lugar a una línea diagonal de 45° , desde la esquina inferior izquierda hasta la superior derecha. La mayoría de las curvas ROC caen entre estos dos extremos. Cualitativamente, cuanto más se acerque una curva ROC a la esquina superior izquierda, más alta será la exactitud global de la prueba. Teniendo en cuenta esto, si se dibujan en un mismo gráfico las curvas ROC obtenidas con distintas pruebas diagnósticas, aquella que esté situada más hacia arriba y hacia la izquierda tendrá mayor exactitud. Por lo tanto, se puede obtener una comparación cualitativa por simple observación de las curvas ROC.

Dadas las ventajas explicadas anteriormente, es fácil de comprender por qué estas curvas se usan habitualmente para tomar decisiones en investigaciones clínicas. Además de proveer información relacionada con cuán frecuentemente las predicciones de un test son correctas, proporcionan un método gráfico para evaluar y discriminar entre diferentes test diagnósticos o incluso modificaciones del mismo test. En particular, cuando es

necesario elegir entre dos test, puede usarse como función discriminante el área debajo de la curva ROC (AUC, Area Under Curve, en inglés). Dicho valor mide la probabilidad de que en pares de sujetos elegidos aleatoriamente (uno de ellos de la población de los casos enfermos y otro de ellos de la población control), el test considerado los clasifique correctamente. Por lo tanto, el AUC sirve como un resumen cuantitativo de la capacidad de asociación subyacente entre el test estadístico y el estado de salud/enfermedad.

A modo de resumen, se listan las ventajas y desventajas que presenta el uso de las curvas ROC en la evaluación de pruebas diagnosticas. Según Zweig y Campbell [60], las principales ventajas son

1. Son una representación fácilmente comprensible de la capacidad de discriminación de la prueba en todo el rango de puntos de corte.
2. Son simples, gráficas y fáciles de interpretar visualmente.
3. No requieren un nivel de decisión particular porque está incluido todo el espectro de puntos de corte.
4. Son independientes de la prevalencia, ya que la sensibilidad y la especificidad se obtienen en distintos subgrupos. Por tanto, no es necesario tener cuidado para obtener muestras con prevalencia representativa de la población. De hecho, es preferible generalmente tener igual número de individuos en ambos subgrupos.
5. Proporcionan una comparación visual directa entre pruebas en una escala común, mientras que otro tipo de gráficos, como los diagramas de puntos o los histogramas de frecuencias, requieren diferentes gráficos cuando difieren las escalas.
6. La especificidad y la sensibilidad son accesibles en el gráfico, en contraste con los diagramas de puntos y los histogramas.

Aparentes desventajas:

1. Los puntos de corte, aunque son conocidos y se han usado para generar el gráfico, no aparecen en él, a diferencia de los diagramas de puntos y los histogramas. Al observar un punto de la curva se observan la sensibilidad y la especificidad que tiene asociadas, pero no se conoce el valor concreto de dicho punto.
2. El número de sujetos de la muestra estudiada tampoco aparece en el gráfico.
3. Al disminuir el tamaño de la muestra, la curva ROC tiende a hacerse más escalonada y desigual. Sin embargo, incluso con gran número de sujetos, la curva puede ser muy desigual.

- La generación de las curvas y el cálculo de sus parámetros son difíciles sin computadora, y los programas existentes no están ampliamente distribuidos.

Como parte del presente trabajo de Tesis y con el fin de agilizar el análisis de los datos, se generó un software específico en el entorno Matlab que analiza las distribuciones y evalúa parámetros de discriminación entre poblaciones. Éste calcula los estadísticos $F, t_{st}, T_w,$ y f , descriptos previamente, y la relación señal/ruido(S/N). La S/N fue calculada como la diferencia de las medias dividida por la varianza combinada, es decir:

$$S/N = \frac{|\bar{X} - \bar{Y}|}{\hat{\sigma}^2}$$

El software, además realiza y grafica automáticamente los histogramas de cada población y las curvas ROC, y de cada una de ellas calcula su AUC. En la figura 4.5 se muestran, a modo de ejemplo, las ventanas que devuelve el software con todos los valores calculados.

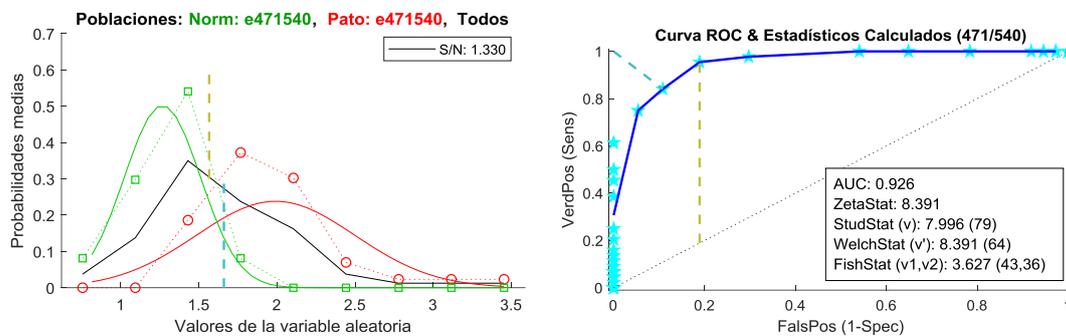


Figura 4.5 Histograma (Izq.) y curva ROC (Der.) representativas.

En el gráfico de la izquierda se presentan los histogramas de cada población y las distribuciones normales estimadas a partir de los respectivos parámetros de posición y dispersión. También se muestra el valor obtenido de la relación señal/ruido,

En la figura de la derecha se muestra la curva ROC "cruda" en estrellas celestes, y la curva ROC "suavizada por intervalos de especificidad", en trazos azules. En la caja de texto de la esquina inferior derecha se muestran el valor del AUC (calculada a partir de la curva azul), y los diferentes estadísticos calculados con sus respectivos grados de libertad. Los estadísticos de medias se calculan siempre con la diferencia en el siguiente orden: "Patológicos - Sanos", de manera que valores positivos indican que la media del índice asociado a los pacientes enfermos es mayor que la media de los pacientes sanos, y viceversa.

Cabe destacar que para verificar el correcto funcionamiento del software desarrollado se hizo un testeó de su rendimiento, corriéndolo con poblaciones simuladas de diferentes características.

4.4 Técnicas numéricas – herramientas de análisis multivariado

Tal como se explicó en el Capítulo 1, las técnicas espectroscópicas son herramientas apropiadas para la realización de estudios no invasivos sobre fluidos corporales que pueden aportar valiosa información en condiciones fisiológicas. En particular, las espectroscopias de absorción UV/VIS y de fluorescencia permiten extraer información acerca los componentes del sistema estudiado. Sin embargo, en los casos que los espectros presentan altos grados de solapamiento espectral, las herramientas tradicionales de interpretación de datos univariados que involucran el análisis de los valores registrados para cada muestra a una única longitud de onda, pueden resultar insuficientes.

En los últimos años se ha extendido el empleo de métodos numéricos para el análisis y la interpretación de conjuntos de datos multivariados. Estas técnicas utilizan modelos “blandos” para obtener información asociada a mezclas complejas. En los Capítulos 5 y 6 del presente trabajo de Tesis se hace uso de diferentes metodologías, basadas en el álgebra matricial y frecuentemente empleadas para el análisis de datos espectroscópicos, para la construcción de modelos blandos que faciliten la interpretación de datos multivariados y permitan el desarrollo de herramientas para el diagnóstico de situaciones patológicas. Entre las metodologías empleadas se encuentran el análisis por componentes principales (PCA), la resolución multivariada de curvas por cuadrados mínimos alternados (MCR-ALS), el análisis discriminante lineal (LDA) y el análisis paralelo de factores (PARAFAC). Dado que estas metodologías serán empleadas para el análisis de la información multidimensional recolectada en los EEm y en las MEE, a continuación se presenta una breve descripción de los fundamentos de cada una de ellas.

4.4.1 Análisis de componentes principales - PCA

En estadística, el análisis de componentes principales (en inglés: PCA) es una técnica de análisis multivariado ampliamente utilizada para reducir la dimensionalidad de un conjunto de datos. [61-62] El objetivo del PCA es transformar una serie de variables posiblemente correlacionadas entre sí en un número menor de variables no correlacionadas denominadas componentes principales (PCs), donde PC1 explica la máxima porción de la varianza de la matriz original, PC2 es ortogonal al primer componente y abarca la mayor cantidad de la variabilidad restante, y así sucesivamente, hasta que se explica el total de la variación de la matriz original por encima del error experimental. Con el PCA se construyen ejes artificiales (PC) que permiten obtener gráficos de dispersión de observaciones y/o variables con propiedades óptimas para la interpretación de la variabilidad y covariabilidad subyacente. Dicho de otra manera, el PCA halla la mejor representación de la variabilidad de los datos en un espacio de menor dimensión.

Si bien el ajuste depende de los datos, en general con unos pocos componentes principales (2 o 3) es suficiente para reducir la dimensionalidad manteniendo la máxima varianza.

La ecuación básica para el PCA puede ser escrita como:

$$\mathbf{X} = \mathbf{T} \mathbf{P}^t + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E} \quad (4.11)$$

donde \mathbf{X} ($I \times J$) es la matriz de datos, con “ I ” muestras, y “ J ” variables. La matriz \mathbf{T} ($I \times R$) contiene los llamados “scores”, que representan la relación entre objetos o muestras y la matriz \mathbf{P}^t ($R \times J$) está formada un conjunto de vectores unitarios (denominados “loadings”) que representan la relación entre las variables a partir de una combinación lineal de ellas. Por último \mathbf{E} ($I \times J$) representa a la matriz de los residuos, es decir la porción de la varianza no explicada por el modelo, y R es el número de componentes principales usados para describir a \mathbf{X} .

De esta manera, en términos matemáticos, PCA es la descomposición de la matriz \mathbf{X} en dos matrices \mathbf{T} y \mathbf{P} de menor dimensión, más una matriz \mathbf{E} de residuos, y $\hat{\mathbf{X}}$ es la estimación que el modelo hace de los valores de \mathbf{X} .

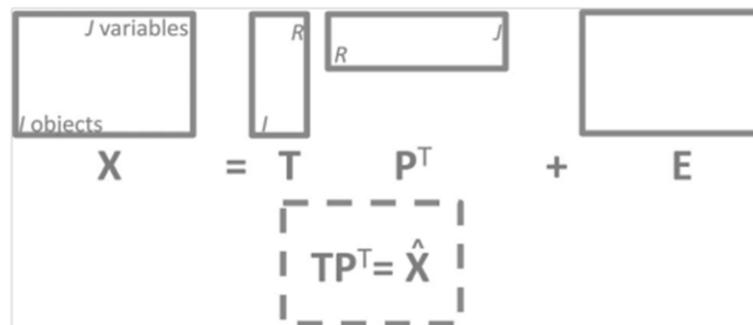


Figura 4.6 La estructura de un modelo PCA.

En el presente trabajo de Tesis se aplicó el PCA a los espectros de emisión de fluorescencia, previamente normalizados a norma euclídeana unitaria. Entonces la matriz \mathbf{X} quedó formada por “ I ” filas, donde cada una de ellas corresponde a una muestra analizada, y “ J ” columnas correspondientes a cada longitud de onda de emisión del espectro de fluorescencia medido.

Para describir el método de manera más detallada, puede pensarse que las variables individuales de la matriz \mathbf{X} (columnas), que se denotan por x_j ($j = 1, \dots, J$), son todos vectores en el espacio I -dimensional. Entonces, una combinación lineal de esas variables se puede escribir como $\mathbf{t} = w_1 x_1 + \dots + w_j x_j$, donde \mathbf{t} es ahora un nuevo vector en el mismo espacio que las variables x (por ser una combinación lineal de estas). En notación matricial, esto se expresa como $\mathbf{t} = \mathbf{X} \cdot \mathbf{w}$, siendo \mathbf{w} el vector con elementos w_j ($j = 1, \dots, J$).

Dado que la matriz \mathbf{X} contiene toda la información, en ella se presenta toda la variabilidad de los datos relevantes para el análisis en cuestión. Entonces, el problema se

puede plantear en términos de conservar la mayor cantidad posible de esa variación también en \mathbf{t} . Si la proporción de la variabilidad capturada en \mathbf{t} es apreciable, entonces puede servir como un buen resumen de las variables \mathbf{x} . Por lo tanto, las J variables de \mathbf{X} pueden ser reemplazadas por una sola variable \mathbf{t} que retenga la mayor parte de la información relevante, (o bien por una cantidad R de variables \mathbf{t} , mucho menor que J , como se verá a continuación).

La variación contenida en \mathbf{t} puede estimarse calculando su varianza. Entonces el problema de reducción de la dimensionalidad consiste en encontrar los pesos óptimos (w_1, w_2, \dots, w_J) que maximicen esa varianza. No obstante, hay que tener en cuenta que los \mathbf{w} necesariamente deben estar normalizados. Esta condición es muy importante para evitar que al multiplicar una w óptima por un número arbitrario “grande” también la varianza de \mathbf{t} sea arbitrariamente grande. Por lo tanto, se les exige a los \mathbf{w} que la suma de sus valores al cuadrado sea igual a uno.

Formalmente el problema se expresa como

$$\underbrace{\text{Máx} (Var(\mathbf{t}))}_{\|\mathbf{w}\|=1} \quad (4.12)$$

y se reduce a encontrar la \mathbf{w} de longitud uno que maximice la varianza de \mathbf{t} . La función “Máx” es la notación matemática para devolver el argumento \mathbf{w} de la función de maximización. Dado que la $Var(\mathbf{t})$ puede escribirse como $\mathbf{t}^T \mathbf{t}$, con $\mathbf{t} = \mathbf{X}\mathbf{w}$, la ecuación anterior puede escribirse en forma más explícita:

$$\underbrace{\text{Máx} (\mathbf{t}^T \mathbf{t})}_{\|\mathbf{w}\|=1} = \underbrace{\text{Máx} (\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w})}_{\|\mathbf{w}\|=1} \quad (4.13)$$

Es importante aclarar aquí que a cada uno de los espectros de \mathbf{X} la rutina PCA resta el espectro promedio de los “ l ” ingresados. De esta manera, lo que en realidad se analiza por PCA, son las diferencias entre el espectro promedio y cada espectro individual a todas las longitudes de onda. Entonces, dado que la matriz \mathbf{X} está centrada en la media, todas las combinaciones lineales previamente mencionada también estarán centradas en la media. El problema planteado en la ecuación 4.13, es un problema estándar en álgebra lineal y la \mathbf{w} óptima es el primer autovector estandarizado, (el autovector con el valor más grande), de la matriz de covarianza $\mathbf{X}^T \mathbf{X} / (n-1)$ o la matriz de productos cruzados correspondiente $(\mathbf{X}^T \mathbf{X})$.

Una manera de cuantificar cuánto refleja \mathbf{t} la información contenida en \mathbf{X} , consiste en proyectar las columnas de \mathbf{X} en \mathbf{t} y calcular los residuos de esa proyección. Esto se realiza empleando la ecuación 4.11, donde \mathbf{p} es el vector de los coeficientes de regresión y \mathbf{E} es la matriz de los residuos. Justamente, \mathbf{p} es igual a \mathbf{w} y todo el mecanismo de regresión puede usarse para evaluar la cantidad de información contenida en \mathbf{t} . Tradicionalmente, esto se hace calculando:

$$\frac{\|X\|^2 - \|E\|^2}{\|X\|^2} \cdot 100\% \quad (4.14)$$

que se conoce como el porcentaje de variación explicada por \mathbf{t} .

Si el porcentaje de variación explicada de la ecuación 4.14 es demasiado pequeño, entonces la combinación \mathbf{t} y \mathbf{p} no es un resumen suficientemente bueno de los datos y se requieren más dimensiones para describir a \mathbf{X} adecuadamente. Entonces, la ecuación 4.11 puede escribirse como:

$$\mathbf{X} = \mathbf{T} \mathbf{P}^t + \mathbf{E} = t_1 \mathbf{p}_1^t + t_2 \mathbf{p}_2^t + \dots + t_R \mathbf{p}_R^t = \hat{\mathbf{X}} + \mathbf{E}, \quad (4.15)$$

donde $\mathbf{T}=[t_1, \dots, t_R](I \times R)$ y $\mathbf{P}=[p_1, \dots, p_R](J \times R)$ ahora son matrices que contienen, respectivamente, R vectores de scores y R vectores de loadings. Si R es mucho más pequeño que J , entonces \mathbf{T} y \mathbf{P} equivalen a una descripción considerablemente más resumida de la variación en \mathbf{X} .

Cabe destacar que cuando se emplea más de una dimensión para representar a \mathbf{X} , el análisis por componentes principales impone como restricción adicional que cada nuevo componente no esté correlacionado con los anteriores. En otras palabras, debe cumplirse simultáneamente que $\mathbf{P}^T \mathbf{P} = \mathbf{I}$ y que $\mathbf{T}^T \mathbf{T}$ sea una matriz diagonal. Esto corresponde a la normalización de los loadings y a la ortogonalidad de los scores mencionadas anteriormente. Cada vector \mathbf{p} , por lo tanto, tiene norma unitaria y es ortogonal a los otros loadings (una base ortonormal). La restricción sobre \mathbf{T} implica que los scores son ortogonales entre sí.

Debido a la ortogonalidad en \mathbf{P} , los R componentes tienen contribuciones independientes a la variación general explicada:

$$\|\mathbf{X}\|^2 = \|t_1 \mathbf{p}_1^t\|^2 + \dots + \|t_R \mathbf{p}_R^t\|^2 + \|\mathbf{E}\|^2 \quad (4.16)$$

y se puede usar la expresión "variación explicada por cada componente", de manera similar a la ecuación 4.14.

Por lo tanto la obtención de componentes principales se resume a obtener los pesos w_{kj} , con k y j variando de 1 a r , del siguiente sistema de r ecuaciones:

$$\begin{aligned} PC_1 &= w_{11} * X_1 + w_{12} * X_2 + \dots + w_{1r} * X_r \\ &\dots \\ PC_r &= w_{r1} * X_1 + w_{r2} * X_2 + \dots + w_{rr} * X_r \end{aligned} \quad (4.17)$$

Donde PC_1 a PC_r : son los componentes principales, que no están correlacionados. Generalmente los PCs se ordenan de acuerdo con la fracción de la información total que cada uno retiene. Es decir que son ordenados en orden decreciente en cuanto a su participación en la varianza de los datos.

Tal como se explicó previamente, los pesos w se estiman de modo que se cumplen las siguientes restricciones:

1) El primer componente, PC_1 , cubre la fracción máxima de la varianza total de los datos del sistema. El segundo componente, PC_2 , cubre la fracción máxima del resto de la varianza total del sistema no incluido en el primer componente y así sucesivamente. De esta forma se obtiene una secuencia de componentes, PC_1 a PC_r , en orden decreciente de importancia (el número máximo de componentes es igual a r).

2) Normalización: $w_{k1}^2 + w_{k2}^2 + \dots + w_{kr}^2 = 1$ con $k=1, \dots, r$. Con esta condición los vectores asociados a los loadings se normalizan. Se adopta para evitar que los valores numéricos de los pesos afecten a la varianza representada por un dado componente en relación a los demás.

3) Ortogonalidad: $w_{k1} * w_{j1} + w_{k2} * w_{j2} + \dots + w_{kr} * w_{jr} = 0$ con $k \neq j$. Esta condición garantiza que los loadings obtenidos no estén correlacionados.

El objetivo de los algoritmos de cálculo de componentes principales es obtener w_{kj} respetando las restricciones 1 a 3. Matemáticamente esto se logra diagonalizando la matriz de covarianza de los datos de modo de obtener sus autovalores y autovectores.

4.4.2 Resolución Multivariada de Curvas por cuadrados mínimos alternados (MCR-ALS)

La Resolución Multivariada de Curvas, MCR (Multivariate Curve Resolution), es la denominación genérica para una familia de métodos cuyo objetivo es la descomposición bilineal de una respuesta instrumental compleja en las contribuciones vinculadas a cada uno de los N componentes puros en el sistema bajo estudio. Para las medidas espectrofluorométricas realizadas en este trabajo de Tesis se asumirá la validez del principio de superposición lineal, es decir que la fluorescencia total a cada longitud de onda es la suma de las contribuciones asociada a la fluorescencia de cada componente del sistema. Cabe destacar que, dada la compleja composición de la orina, la expresión "componente del sistema" se referirá a un concepto más bien general que no necesariamente se corresponde con el espectro de fluorescencia de una única sustancia presente en la misma. Es decir, cada componente corresponde a una contribución espectral asociada a la presencia, en determinadas condiciones, de un conjunto de fluoróforos que no se intentan identificar individualmente.

La matriz de datos a analizar (D), está formada por el conjunto de los espectros de emisión de fluorescencia medidos a partir de la orina de diferentes personas. El objetivo del MCR-ALS es descomponer a la matriz D en otras dos matrices C y S correspondientes a

los scores y a los perfiles espectrales respectivamente. A continuación se expresa esta descomposición bilineal en forma matricial:

$$D = CS^T + E \quad (4.18)$$

donde la matriz **E** contiene la porción de la variación no explicada por el modelo (es decir los residuos). La ecuación 4.18 puede representarse esquemáticamente como:

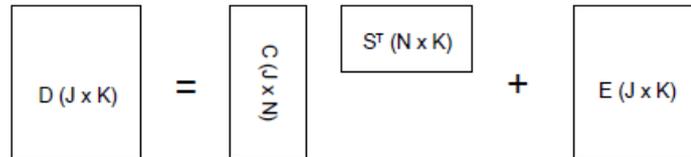


Figura 4.7 Esquema de la descomposición matricial de la matriz *D*.

Matemáticamente para las muestras bajo estudio se debe cumplir que:

$$d_{jk} = \sum c_{jn} s_{nk} + e_{jk} \quad (4.19)$$

donde d_{jk} corresponde a la fluorescencia de la muestra *j* para una longitud de onda *k*, c_{jn} es score del componente *n*-ésimo para esa muestra, s_{nk} es la intensidad relativa de emisión de fluorescencia a la longitud de onda *k* del componente *n*-ésimo y el e_{jk} es el valor residual en dicho punto. El método MCR-ALS utiliza la Ec. 4.18 en forma iterativa, aplicando el algoritmo de mínimos cuadrados alternos ALS (Alternating Least Squares) con restricciones que permiten eliminar de los scores y de los perfiles espectrales las contribuciones sin significado físico. La convergencia del algoritmo ALS se consigue cuando en dos ciclos consecutivos la diferencia de los residuos entre la iteración anterior y la nueva es menor a un valor porcentual predeterminado.

Para llevar a cabo un análisis MCR-ALS es necesario seguir una serie de pasos que se representan esquemáticamente en la figura 2.13 y se describen a continuación:

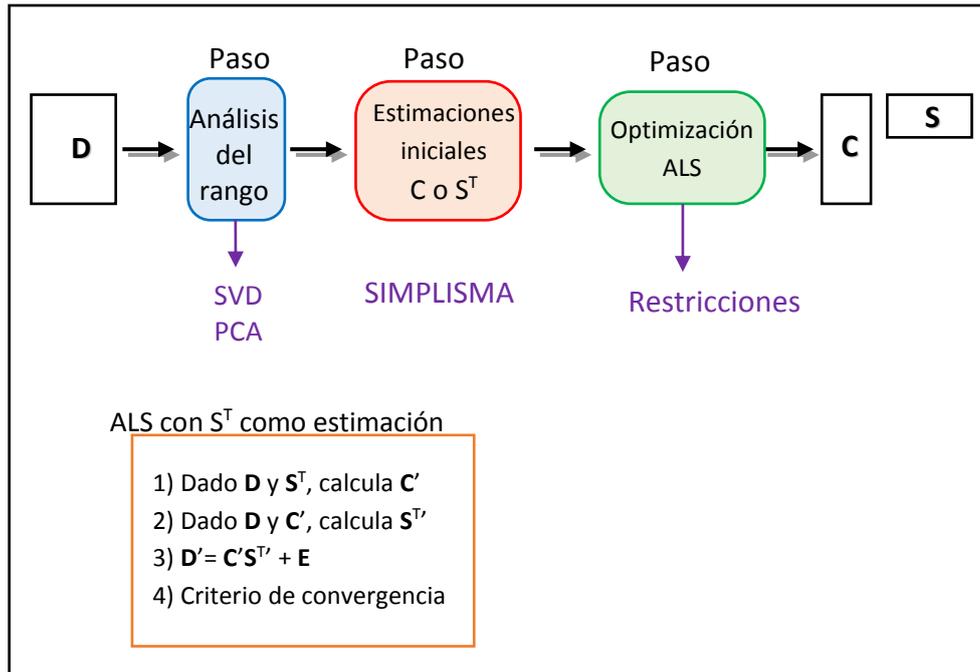


Figura 4.8 Pasos en una descomposición MCR-ALS de la matriz D ($J \times K$).

Paso 1: Estimación del número de componentes:

Para la estimación del número de componentes independientes existen diversas metodologías, entre las más empleadas podemos nombrar el análisis de componentes principales (PCA) y la descomposición en valores singulares (SVD). Tal como se explicó en la sección anterior, el PCA es un método bilineal que permite reproducir la matriz original con un número menor de variables no correlacionadas entre sí, para reducir la dimensionalidad. La SVD [63] es una generalización del análisis de autovalores y autovectores para matrices rectangulares (no cuadradas). Consiste en descomponer la matriz D ($J \times K$) en tres matrices U ($J \times J$), S ($J \times K$) y V ($K \times K$) de la forma $D=USV^T$, donde la matriz S es una matriz diagonal que contiene a los valores singulares, y las matrices U y V , corresponden a los vectores singulares “por izquierda” y “por derecha” respectivamente, que sirven como bases para el espacio de las filas y las columnas de D . Es decir, la matriz U ($J \times J$), que mapea las filas, es una base para dicho espacio, mientras que la matriz V ($K \times K$) es una base para el espacio de las columnas. Entonces, cualquier vector fila puede ser expresado como una combinación lineal de los versores de U , mientras que cualquier vector columna puede ser expresado como una combinación lineal de los versores de V . Tal como en el caso del PCA, los valores singulares se ordenan de mayor a menor, siendo los primeros los que permiten una mejor representación de la matriz D .

Paso 2: Necesidad de estimaciones iniciales

La resolución de curvas por mínimos cuadrados alternos requiere de estimaciones iniciales ya sea de los scores o de los perfiles espectrales antes de iniciar la optimización

del algoritmo. Debido a que una buena estimación inicial favorece la convergencia del algoritmo hacia la solución real, para los sistemas estudiados en este trabajo de tesis se utilizó la metodología matemática SIMPLISMA. Este procedimiento consiste, en primer lugar, en hallar la fila con la mayor desviación estándar relativa. Seguidamente, se procede a la normalización de dicha fila. La segunda variable de mayor “pureza”, además de poseer la mayor desviación estándar, entre las filas restantes, deberá presentar la mínima correlación con la primera variable “pura” encontrada. Para seleccionar una nueva variable, se calcula el determinante de la matriz compuesta por las variables puras halladas y la fila de la matriz de datos que aún no ha sido elegida. El valor calculado para el determinante será proporcional a la independencia que haya entre las variables puras halladas y la fila seleccionada. Así, el determinante tendrá un valor más alto en la medida que las variables no estén correlacionadas, mientras que, por el contrario para variables correlacionadas dará un valor cercano a cero. El procedimiento se repite hasta encontrar tantos vectores no correlacionados como componentes se hayan estimado en el paso 1.

Paso 3: Optimización

Como se mencionó anteriormente, MCR utiliza la ecuación 4.18 en forma iterativa, aplicando el algoritmo ALS. A modo de ejemplo, conocida la matriz \mathbf{D} y estimada la matriz de espectros \mathbf{S}^T , se desea encontrar una matriz \mathbf{C} que minimice la suma de cuadrados de los elementos de la matriz de residuos ($\mathbf{E}=\mathbf{D}-\mathbf{C}*\mathbf{S}^T$). Dado que en general, la matriz \mathbf{S} no es cuadrada, el cálculo de la matriz \mathbf{C} se realiza por cuadrados mínimos. Esto se logra empleando el concepto de matriz pseudoinversa que se explica a continuación.

Partiendo de la descomposición de \mathbf{D} ($\mathbf{D}=\mathbf{C}*\mathbf{S}^T$) y multiplicando “por derecha” por la matriz \mathbf{S} en ambos miembros, se obtiene:

$$\mathbf{D}*\mathbf{S}=\mathbf{C}*(\mathbf{S}^T*\mathbf{S}) \quad (4.20)$$

La matriz $\mathbf{S}^T*\mathbf{S}$, es cuadrada y por tanto admite inversa. Multiplicando nuevamente por derecha por la inversa de $\mathbf{S}^T*\mathbf{S}$ se llega a la ecuación 4.21.

$$\mathbf{D}*\mathbf{S}*(\mathbf{S}^T*\mathbf{S})^{-1}=\mathbf{C}*(\mathbf{S}^T*\mathbf{S})*(\mathbf{S}^T*\mathbf{S})^{-1} \quad (4.21)$$

donde $(\mathbf{S}^T*\mathbf{S})*(\mathbf{S}^T*\mathbf{S})^{-1}$ es igual a la matriz identidad, mientras que la expresión $\mathbf{S}*(\mathbf{S}^T*\mathbf{S})^{-1}$ es denominada pseudoinversa de la matriz \mathbf{S}^T y se representa como $(\mathbf{S}^T)^+$. De esta manera, la ecuación 4.18 puede escribirse como:

$$\mathbf{D}*(\mathbf{S}^T)^+=\mathbf{C} \quad (4.22)$$

Puede demostrarse que la matriz \mathbf{C} estimada de esta forma corresponde al ajuste óptimo (por cuadrados mínimos) del sistema de ecuaciones lineales definido por la ecuación 4.19 que, al estar sobredeterminado, no posee solución única.

Aplicando el mismo razonamiento, pero partiendo de la estimación de C , se obtiene una estimación por cuadrados mínimos para la matriz S^T

$$(C)^+ * D = S^T \quad (4.23)$$

De esta manera el proceso iterativo se realiza aplicando alternadamente las ecuaciones 4.22 y 4.23.

Cabe destacar que, aun habiendo hecho un buen análisis del número de componentes y teniendo buenas estimaciones iniciales, la factorización de la matriz D puede arrojar vectores sin significado físico debido a la existencia de ambigüedades. Las ambigüedades matemáticas implican que la resolución de la matriz D puede conseguirse, sin perder calidad de ajuste, con infinitos pares de matrices C y S^T diferentes. Para minimizar el efecto de las mismas se utilizan restricciones que permiten darle un sentido físico/químico a la solución obtenida. Las restricciones utilizadas en el presente trabajo de Tesis fueron:

- **No-negatividad**: Se obliga a que tanto los scores como los perfiles espectrales sean siempre positivos. Esta restricción es aplicable a resultados espectrofluorométricos en donde las contribuciones a la fluorescencia total de los componentes de la mezcla son siempre positivas.

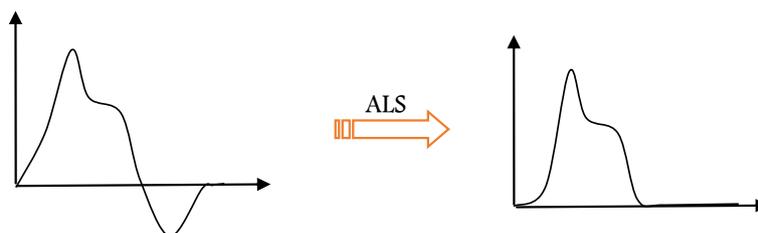


Figura 4.9 Restricción de no negatividad.

- **Unimodalidad**: esta restricción se puede aplicar a aquellos perfiles en los que se presupone que solo poseen un máximo, por lo cual este tipo de restricción suele utilizarse para señales que posean forma de pico o banda. En particular los espectros de emisión correspondientes a un único estado emisor deben cumplir con esta condición.

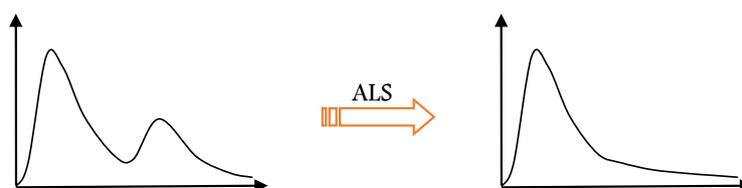


Figura 4.10 Restricción de unimodalidad.

Una vez concluida la resolución de la matriz **D**, la calidad de ajuste del modelo MCR-ALS puede ser evaluada mediante el porcentaje de varianza explicada (R^2), la desviación estándar de los residuos (σ) y la falta de ajuste de los datos (LOF). Esta última se define como la diferencia entre los datos de entrada **D** (ya sea la matriz de datos experimentales crudos o la matriz de datos reproducida por el PCA en el ciclo iterativo anterior) y los datos reproducidos a partir del producto obtenido por \mathbf{CS}^T .

Las ecuaciones que definen estos parámetros son:

$$R^2 (\%) = \left(\frac{\sum_{i,j} d_{ij}^2 - \sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2} \right) * 100 \quad (4.24)$$

$$\sigma = \sqrt{\frac{\sum_{i,j} e_{ij}^2}{n_r n_c}} \quad (4.25)$$

$$LOF (\%) = \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} * 100 \quad (4.26)$$

donde d_{ij} es un elemento de la matriz de datos experimentales, e_{ij} es el valor residual obtenido de la diferencia entre los datos experimentales (matriz **D**) y los datos reproducidos (producto \mathbf{CS}^T de las matrices obtenidas por MCR-ALS), n_r y n_c son el número de filas y de columnas de la matriz **D**.

Software empleado

Los espectros de fluorescencia fueron exportados y convertidos al entorno de trabajo de MATLAB y posteriormente se analizaron con la rutina MCR-ALS Toolbox 1.0 en MATLAB R2008a.

4.4.3 Análisis discriminante lineal (LDA)

El Análisis Discriminante Lineal (en inglés, LDA) es una generalización del discriminante lineal de Fisher, un método utilizado en estadística para el reconocimiento de patrones a partir de una combinación lineal de rasgos que caracterizan o separan dos o más clases de objetos o eventos. De modo similar al PCA, el LDA apunta a una reducción dimensional. Sin embargo, la diferencia entre ambos radica en su objetivo final: PCA se usa para reducir la dimensionalidad de los datos manteniendo la máxima varianza, mientras que LDA intenta maximizar la separabilidad entre las clases de datos presentes. [64] Esto se consigue maximizando la relación de “la varianza entre las diferentes clases” respecto de “las varianzas dentro de cada clase”. Con esta metodología se logra, entonces, que las

muestras de la misma clase sean proyectadas cerca unas de otras pero al mismo tiempo las medias proyectadas para cada clase estén lo más alejadas posible entre sí. La figura 4.11 muestra un esquema de cómo serían las reducciones dimensionales asociadas al PCA y al LDA, para el mismo grupo de datos. En la figura el término "representación" se refiere a capturar la mayor variabilidad de los datos de entrada, mientras que "clasificación" se refiere a optimizar la discriminación entre clases.

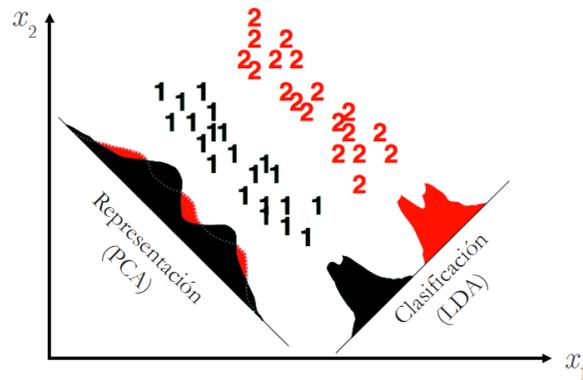


Figura 4.11 Esquema representativo de la reducción dimensional por PCA y por LDA.

El fundamento matemático del LDA se describe a continuación:

Considérese un conjunto de N observaciones, $\mathbf{X}=\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, donde cada muestra, $\mathbf{x}_k=[x_{k1}, \dots, x_{kD}]^T$, es un vector D -dimensional que pertenece a la clase w_i , con $i=1, 2, \dots, C$, donde " C " es el número de clases. Suponiendo, tal como es el caso del presente trabajo de Tesis, que sólo existen 2 clases posibles ($C=2$), entonces los patrones tratan de proyectarse sobre una recta. Ese es el caso que se desarrolla a continuación, considerando, además, que N_1 muestras pertenecen a la primera clase (Clase 1), y N_2 pertenecen a la segunda clase (Clase 2), siendo $N_1+N_2=N$. LDA transforma de $\mathbf{X} \rightarrow \mathbf{Y}$, con $\mathbf{Y}=\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$, a través de la proyección de las muestras \mathbf{X} sobre un hiperplano con dimensión $C-1$. Por lo tanto, reduce la dimensionalidad de los datos de \mathbb{R}^D a \mathbb{R}^{C-1} .

Consideraremos que la línea que discrimina entre clases está dada por el vector unitario \mathbf{v} . Entonces, el escalar $\mathbf{v}^T \mathbf{x}_k$ es la distancia de la proyección de \mathbf{x}_k al origen (ver figura 4.12).

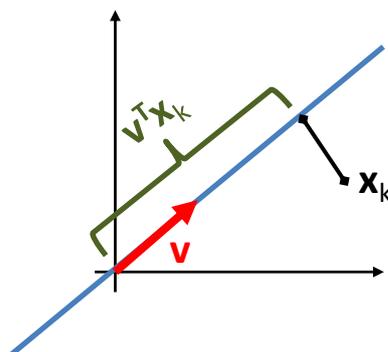


Figura 4.12 Esquema representativo donde se muestran las direcciones de los vectores para el caso de un plano.

En otras palabras, la proyección de \mathbf{x}_k en un espacio unidimensional definido por \mathbf{v} está dada por

$$y_k = \mathbf{v}^T \mathbf{x}_k \quad (4.27)$$

y de esta manera, la función de regresión lineal en el espacio 2D $f(\mathbf{x}_k; \mathbf{v}) = \mathbf{v}^T \mathbf{x}_k$ proyecta a cada punto del conjunto de muestras \mathbf{X}_k (de \mathbb{R}^2), sobre la recta que pasa por el origen y es paralela al versor \mathbf{v} .

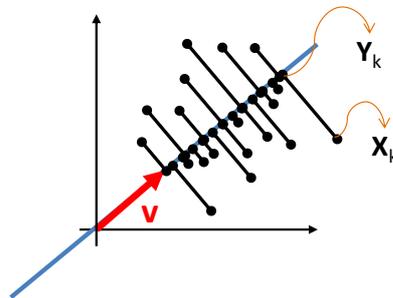


Figura 4.13 Esquema representativo donde se muestran \mathbf{X}_k e \mathbf{Y}_k , junto con las direcciones de los vectores para el caso de un plano.

Es fácil inferir, entonces, que al variar el vector \mathbf{v} se obtienen diferentes niveles de separación entre las clases generados por la proyección de los puntos originados sobre la recta discriminante. Por lo tanto, la solución al problema se reduce a encontrar el vector \mathbf{v} que maximice la separación entre las clases proyectadas, de manera de optimizar la proyección.

Para cuantificar la separación entre clases, se calcula la distancia entre las medias de cada clase, en el espacio proyectado, que está dada por la ecuación 4.28.

$$D = (\tilde{\mathbf{m}}_1 - \tilde{\mathbf{m}}_2)^2 \quad (4.28)$$

donde los vectores de medias para la clase w_i ($i=1,2$), en los espacios original y proyectado están dadas por las ecuaciones 4.29 y 4.30, respectivamente.

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \mathbf{x}_k \quad (4.29)$$

$$\tilde{\mathbf{m}}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} y_k = \frac{1}{N_i} \sum_{k=1}^{N_i} \mathbf{v}^T \mathbf{x}_k = \mathbf{v}^T \mathbf{m}_i \quad (4.30)$$

Por lo tanto, expresando la ecuación 4.28 en términos del espacio original la distancia entre medias resulta $D = |\mathbf{v}^T \mathbf{m}_1 - \mathbf{v}^T \mathbf{m}_2|$.

Cabe destacar que cuanto mayor sea \mathbf{D} , más distantes estarán las medias de ambas clases. Sin embargo, esta función no tiene en cuenta la variabilidad de los datos intra-clase, es decir, no tiene en cuenta la dispersión de los datos de cada clase respecto de su media. Para visualizar esta idea, en la figura 4.14 se muestran dos proyecciones posibles (según los ejes X_1 y X_2), del mismo conjunto de datos.

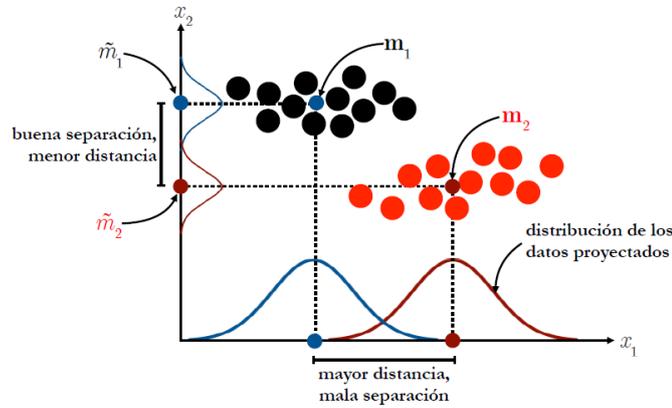


Figura 4.14 La función $D(\mathbf{v})$ basada en la distancia entre clases no tiene en cuenta la dispersión de los datos dentro de cada clase.

Allí puede verse que la proyección de los datos sobre el eje X_1 , presenta una mayor distancia entre las medias de cada clase, respecto de la proyección sobre eje X_2 . Sin embargo, el criterio de separación no es bueno, dado que entre las medias asociadas al eje X_1 existe una zona de superposición de las proyecciones de los datos de ambas clases. Por otro lado, si bien la distancia entre medias en la proyección sobre el eje X_2 es mucho menor, la clasificación en clases mejora notablemente. De aquí que para optimizar la clasificación sea necesario, además, considerar la variabilidad de los datos dentro de cada clase. Dicha variabilidad “intra-clase” se calcula, para cada clase i ($i=1,2$), según la ecuación 4.31, donde la varianza proyectada de la clase w_i está dada por \widetilde{S}_i^2 .

$$\widetilde{S}_i^2 = \frac{1}{N_i} \sum_{k=1}^{N_i} (y_k - \widetilde{m}_i)^2 \quad (4.31)$$

De aquí que un valor \widetilde{S}_i^2 pequeño indica que las muestras proyectadas de la clase i se encuentran muy agrupadas alrededor de la media proyectada de la clase i , mientras que un valor grande implica lo contrario.

En consecuencia optimizar la clasificación es equivalente a encontrar la dirección del vector \mathbf{v} sobre la que los datos proyectados maximicen $J(\mathbf{v})$:

$$J(\mathbf{v}) = \frac{(\widetilde{m}_1 - \widetilde{m}_2)^2}{\widetilde{S}_1^2 + \widetilde{S}_2^2} \quad (4.32)$$

Para resolver esta ecuación, es necesario reescribir $J(\mathbf{v})$ de manera que quede explícitamente en función de \mathbf{v} , tal como se explica a continuación.

La matriz de dispersión “intra-clase” en el espacio original está dada por la ecuación 4.33, y se calcula a partir de la matriz de covarianza de cada clase (4.34)

$$S_w = S_1 + S_2 \quad (4.33)$$

$$S_i = \frac{1}{N_i} \sum_{k=1}^{N_i} (x_k - m_i)(x_k - m_i)^T \quad (4.34)$$

Usando la ecuación 4.30 y la dispersión de los datos proyectados (dada por la ecuación 4.31), se puede reescribir a \widetilde{S}_i^2 en función de la matriz de dispersión en el espacio original como:

$$\begin{aligned} \widetilde{S}_i^2 &= \frac{1}{N_i} \sum_{k=1}^{N_i} (y_k - \widetilde{\mathbf{m}}_i)^2 = \frac{1}{N_i} \sum_{k=1}^{N_i} (v^T x_k - v^T m_i)^2 = \\ &= \frac{1}{N_i} \sum_{k=1}^{N_i} (v^T (x_k - m_i)(x_k - m_i)^T v) = v^T S_i v \end{aligned} \quad (4.35)$$

por lo tanto:

$$\widetilde{S}_1^2 + \widetilde{S}_2^2 = v^T S_w v \quad (4.36)$$

De manera similar, la diferencia entre las medias proyectadas se puede expresar en términos del espacio original:

$$(\widetilde{\mathbf{m}}_1 - \widetilde{\mathbf{m}}_2)^2 = (v^T m_1 - v^T m_2)^2 = v^T (m_1 - m_2)(m_1 - m_2)^T v = v^T S_B v \quad (4.37)$$

donde S_B es la matriz de dispersión interclase.

Finalmente, el discriminante lineal de Fisher en términos de S_w y S_b se define como:

$$J(v) = \frac{v^T S_B v}{v^T S_w v} \quad (4.38)$$

Para optimizar $J(v)$ se deriva con respecto a v y se iguala a cero:

$$\frac{\partial J(v)}{\partial v} = S_B v - J(v) S_w v = 0 \quad (4.39)$$

Y resolviendo el problema de valores propios generalizado

$$S_B v = \lambda S_w v \quad (4.40)$$

Si la matriz S_w admite inversa, entonces la ecuación 4.40 se puede convertir en un problema de valor propio estándar:

$$S_w^{-1} S_B v = \lambda v \quad (4.41)$$

En el caso particular de dos clases, no es necesario resolver los autovectores y autovalores de $S_w^{-1} S_B$ debido a que $S_B v$ siempre tiene la misma pendiente que el vector $\mathbf{m}_1 - \mathbf{m}_2$. Por lo tanto, dado que el factor de escala λ para v no es importante (sino sólo su orientación), se puede encontrar una solución inmediata para v que maximice la función $J(v)$:

$$v = S_w^{-1}(m_1 - m_2) \quad (4.42)$$

De este modo, se obtiene \mathbf{v} para el discriminante lineal de Fisher, que logra la máxima relación entre la dispersión “inter-clase” respecto de la dispersión “intra-clase”. La figura 4.15 muestra un esquema representativo donde puede observarse el versor que define la línea que mejor clasifica los datos de la figura 4.14.

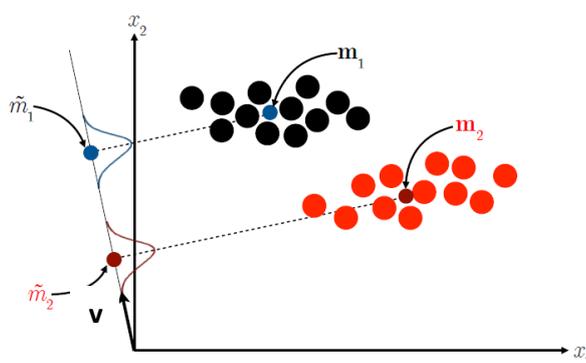


Figura 4.15 Clasificación según LDA, tiene en cuenta la dispersión de los datos entre clases y dentro de cada clase.

4.4.4 Análisis paralelo de factores (PARAFAC)

El análisis paralelo de factores (PARAFAC) es un método de descomposición que conceptualmente puede considerarse como una generalización del análisis de componentes principales (PCA) para un mayor número de dimensiones [65]. Se trata de un método robusto y muy eficiente, que tiene la capacidad de procesar un gran número de muestras. Un modelo PARAFAC descompone a un arreglo tridimensional (\mathbf{X}) en tres matrices bidimensionales \mathbf{A} ($I \times F$), \mathbf{B} ($J \times F$) y \mathbf{C} ($K \times F$), con elementos a_{if} , b_{jf} y c_{kf} cuyos índices i , j y k corresponden a las dimensiones del arreglo \mathbf{X} y f corresponde a los factores empleados para la descomposición. En la Figura 4.16, cada color (rojo, azul y verde) representa una de las dimensiones del arreglo tridimensional y \mathbf{E} contiene la fracción de la varianza no explicada por el modelo [66]. Dado que la descomposición de los datos se realiza en componentes trilineales, el algoritmo se ha aplicado frecuentemente para analizar datos espectrofluorométricos, en particular a conjuntos de matrices de excitación-emisión (MEE) ordenados en arreglos de 3 vías.

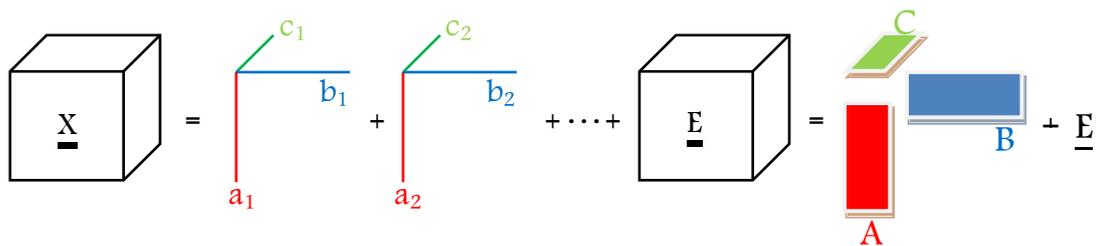


Figura 4.16. Descomposición PARAFAC de datos con estructuras de tres vías.

En este contexto, las matrices **A**, **B** y **C** contienen los perfiles de concentración relativa, los espectros de emisión y los espectros de excitación correspondientes a cada uno de los factores que contribuyen a la fluorescencia en las muestras analizadas. Cabe destacar que, en la descomposición PARAFAC, cada factor viene definido por tres vectores ($a[:,f]$, $b[:,f]$ y $c[:,f]$) los cuales describen su comportamiento en cada una de las tres dimensiones. En el caso de arreglos construidos a partir de MEE los índices i , j y k , están asociados a las muestras, las longitudes de onda de emisión y las longitudes de onda de excitación, respectivamente.

Matemáticamente, el modelo PARAFAC supone que cada elemento del arreglo de 3 vías **X** puede representarse como la superposición de las contribuciones correspondientes a cada factor:

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk}$$

Una de las principales ventajas de este modelo es, en muchos casos, la unicidad de su solución. La solución se obtiene mediante optimización por mínimos cuadrados alternados y durante el proceso iterativo se pueden imponer restricciones a las soluciones a fin obtener perfiles espectrales con sentido físico o que se ajusten mejor a las características físicoquímicas del sistema estudiado. Cabe destacar que frecuentemente los arreglos de datos obtenidos mediante la medición de MEE para diferentes condiciones experimentales obedecen a una estructura trilineal. En estos casos si el número de factores se elige correctamente, el algoritmo PARAFAC es capaz de extraer sin ambigüedades tanto los scores como los perfiles espectrales (de excitación y de emisión), aún para sistemas con alto grado de solapamiento.

El algoritmo general PARAFAC-ALS fue descrito por R. Bro en 1997 [66] y podría esquematizarse de la siguiente manera:

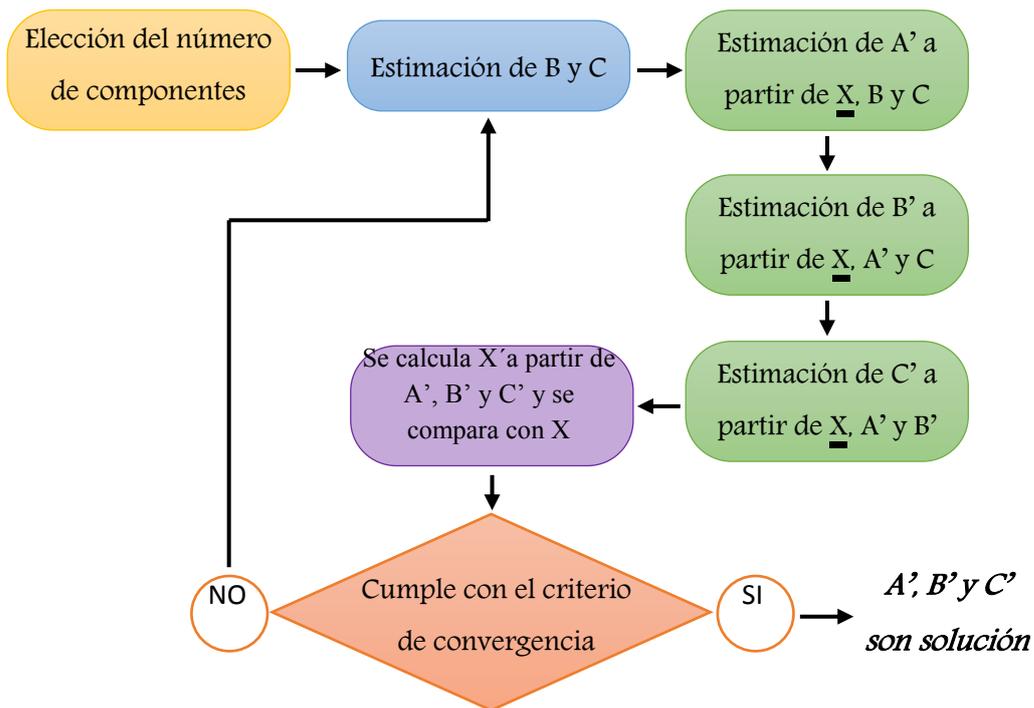


Figura 4.17 Algoritmo general PARAFAC-ALS

A diferencia de otros modelos (por ejemplo, PCA) PARAFAC no se puede aplicar de forma secuencial y es necesario reestimar todo el modelo si se quiere probar un nuevo número de factores.

Tal como se muestra en la Figura 4.17, durante el proceso de optimización, PARAFAC aplica alternativamente diferentes “desdoblamientos” del arreglo de 3 vías para obtener relaciones entre arreglos de 2 vías (matrices) cuya solución se obtiene mediante el algoritmo ALS. Es decir que a lo largo del proceso iterativo, X es desdoblada varias veces de acuerdo con la dimensión que se desea estimar. Por ejemplo X puede desdoblarse en una matriz X ($I \times JK$), lo que implica que los datos que inicialmente eran de tres vías se convierten en datos de dos vías.

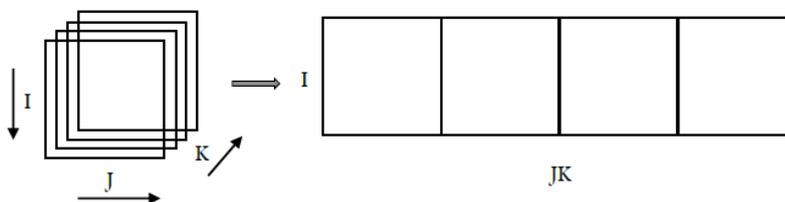


Figura 4.18 Desdoblamiento de la matriz X

Entonces la estimación de A (representada por A') se lleva a cabo resolviendo por mínimos cuadrados la siguiente ecuación

$$X = A'Z^T$$

donde \mathbf{Z} ($F \times JK$) es la matriz calculada a partir de las estimaciones de \mathbf{B} y \mathbf{C} .

Como las matrices no son cuadradas, para realizar esta operación se aplica el mismo razonamiento que el descrito por las ecuaciones 4.20 y 4.21, obteniendo la expresión:

$$A' = X(Z^T)^+$$

donde $(Z^T)^+$ corresponde a la pseudoinversa de Z^T .

De la misma forma para obtener \mathbf{B}' , \mathbf{X} se desdobla en una matriz \mathbf{X} ($J \times IK$) y \mathbf{Z} ($F \times IK$) es calculada a partir de \mathbf{A}' y \mathbf{C} . De manera similar se procede con \mathbf{C}' .

En cada paso de iteración, la calidad de ajuste del modelo mejora y, mediante la aplicación de restricciones, se puede forzar a la solución del algoritmo PARAFAC para obtener un modelo fisicoquímicamente interpretable. Una desventaja que presenta este tipo de análisis es el tiempo requerido para estimar los modelos, ya que a veces son necesarios cientos o miles de iteraciones antes de alcanzar la convergencia. Sin embargo, una buena estimación de valores iniciales podría, en principio, acelerar el algoritmo y asegurar la convergencia. Han sido propuestos varios tipos de iniciación, como por ejemplo empezar desde valores iniciales aleatorios, o iniciar el algoritmo de distintos puntos de partida. En algunos trabajos se propone usar valores iniciales basados en la descomposición en valores propios [67, 68].

PARAFAC aplicado a datos de fluorescencia.

Las matrices de Excitación-Emisión (MEEs) se obtienen combinando espectros de emisión de fluorescencia medidos para un rango de longitudes de onda de excitación. De esta forma, los datos se adaptan a un modelo trilineal, siendo los vectores asociados a cada factor descriptores de los respectivos perfiles de concentración, emisión y excitación (Figura 4.19). [66]

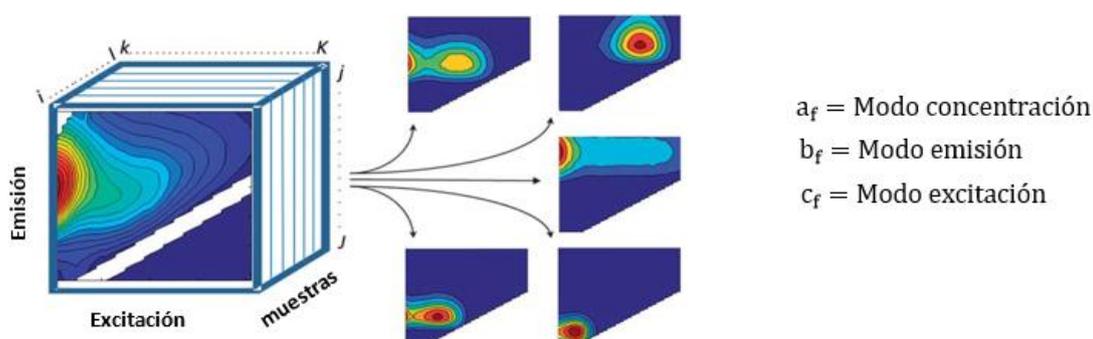


Figura 4.19. Conjunto de datos compuestos por MEEs organizados en una estructura tridimensional descompuesto en 5 componentes mediante PARAFAC. Adaptada de [69].

Sin embargo, la construcción de un modelo PARAFAC para un conjunto de MEEs no es algo trivial e involucra una serie de etapas que incluyen: el pre-procesamiento de las señales registradas, la exploración del conjunto de datos, el desarrollo del modelo y la validación del mismo.

Software empleado.

Las MEE fueron exportadas y convertidas al entorno de trabajo de MATLAB 2018a. El pre-procesamiento de las señales registradas y el análisis de los modelos obtenidos se realizó con un software preexistente al presente trabajo de Tesis, que fue desarrollado por el grupo de trabajo dirigido por el Dr. García Einshlag. Los modelos se obtuvieron empleando la rutina PARAFAC de acceso público [69].

Pre-procesamiento:

Tal como se explicó previamente, el modelo PARAFAC se basa en la descomposición trilineal de las intensidades de emisión registradas en las MEE para diferentes muestras. [66] Por lo tanto, antes de aplicar el algoritmo es necesario realizar un pre-procesamiento de los datos para eliminar contribuciones no lineales presentes en las señales registradas. Además en este paso se llevan a cabo una serie de correcciones que tienen como objetivo atenuar cualquier tipo de sesgo sistemático en el conjunto de datos, eliminar señales que no están relacionadas con la fluorescencia y normalizar el conjunto de datos cuando las diferencias de fluorescencia entre diferentes muestras son muy pronunciadas[69].

En este trabajo de Tesis se aplicaron las siguientes etapas de pretratamiento:

i) Corrección de los efectos de filtro interno

Tal como se explicó en la sección 4.1.2, los efectos de filtro interno ocurren cuando la radiación, incidente y/o emitida, es absorbida por las especies presentes en la muestra, reduciendo la cantidad de radiación efectiva que llega al detector. En el presente trabajo las MEEs han sido corregidas mediante el empleo de sus espectros de absorción, usando la siguiente ecuación:

$$I_{Corr} = I_{Reg} \times 10^{\frac{Abs_{Em} + Abs_{Ex}}{2}}$$

en la que I_{Corr} es la intensidad de emisión corregida, I_{Reg} es la intensidad de emisión registrada, Abs_{Em} corresponde la absorbancia de la muestra a la longitud de onda de emisión y Abs_{Ex} corresponde la absorbancia de la muestra a la longitud de onda de excitación. Calculando estos factores se construyó una matriz de factores de corrección para cada una de las muestras (Figura 4.20), de manera de que cada MEE se multiplicó por su correspondiente matriz de corrección.

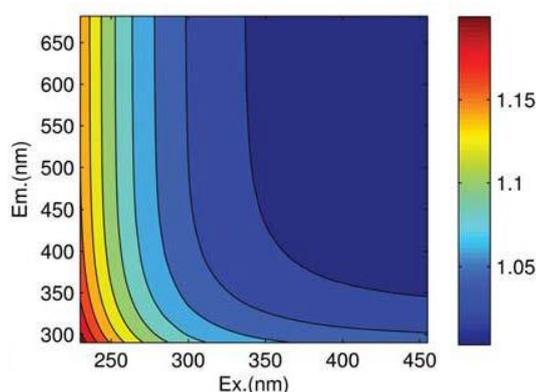


Figura 4.20 Ejemplo de una matriz de factores de corrección calculada a partir de la absorbancia de una muestra de materia orgánica disuelta. [69]

ii) Corrección por el blanco a partir de la sustracción de la matriz correspondiente a las emisiones del solvente

Para realizar esta corrección se utilizó agua bidestilada como blanco. La matriz MEE del blanco se midió cada día de trabajo, y a cada MEE de orina le corresponde el blanco medido ese mismo día. De esta manera, restar la matriz del agua además de servir para eliminar la contribución de luz espuria⁷ y la dispersión Raman del agua, sirve para equiparar las diferencias de emisión total que pueda presentar la lámpara del espectrofluorómetro distintos días.

iii) Eliminación de las regiones de las MEE gobernadas por efectos de dispersión de la radiación tanto de tipo elástico (Rayleigh) como inelástico (Raman).

Dado que los fenómenos de dispersión no pueden ser modelados empleando ecuaciones lineales, las regiones o “calles” correspondientes a dichos fenómenos deben ser o bien eliminadas del conjunto de datos o bien “corregidas”. En las matrices de fluorescencia se observan dos dispersiones características: la de Rayleigh y la de Raman (ambas de primer y segundo orden). La dispersión de Rayleigh de primer orden aparece cuando la longitud de onda de excitación es igual a la de emisión y la de segundo orden cuando la longitud de onda de la emisión es el doble que la de excitación. En muchos casos la magnitud de esta dispersión supera incluso a la señal de fluorescencia. Las dispersiones de Raman, en cambio, suelen ser de menor intensidad, aparecen a longitudes de onda

⁷ La luz espuria es toda aquella luz que llega al detector, pero su origen no se corresponde con el fenómeno físico que se quiere detectar. Puede ser producto de reflexiones internas en el equipo, con la celda, o puede llegar al detector sin haber atravesado la muestra por ejemplo.

ligeramente superiores a las anteriores y su contribución es prácticamente eliminada, al restar la matriz del solvente.

Para los casos en los que las calles de dispersión son eliminadas, el algoritmo PARAFAC permite manejar arreglos de 3 vías con elementos faltantes (que en el entorno de Matlab se indican como NaN, Not a Number). Por otro lado, en muchas situaciones, para acelerar la convergencia del algoritmo, los elementos del arreglo de 3 vías ubicados en las posiciones correspondientes a las “calles” de dispersión son reemplazados por valores calculados mediante interpolación cúbica segmentaria (“Cubic Spline” del toolbox “nway” de Matlab) empleando datos de las regiones adyacentes a las “calles” de dispersión. [70] En la figura 4.21 se muestran esquemáticamente los efectos de los diferentes pasos de pre-tratamiento.

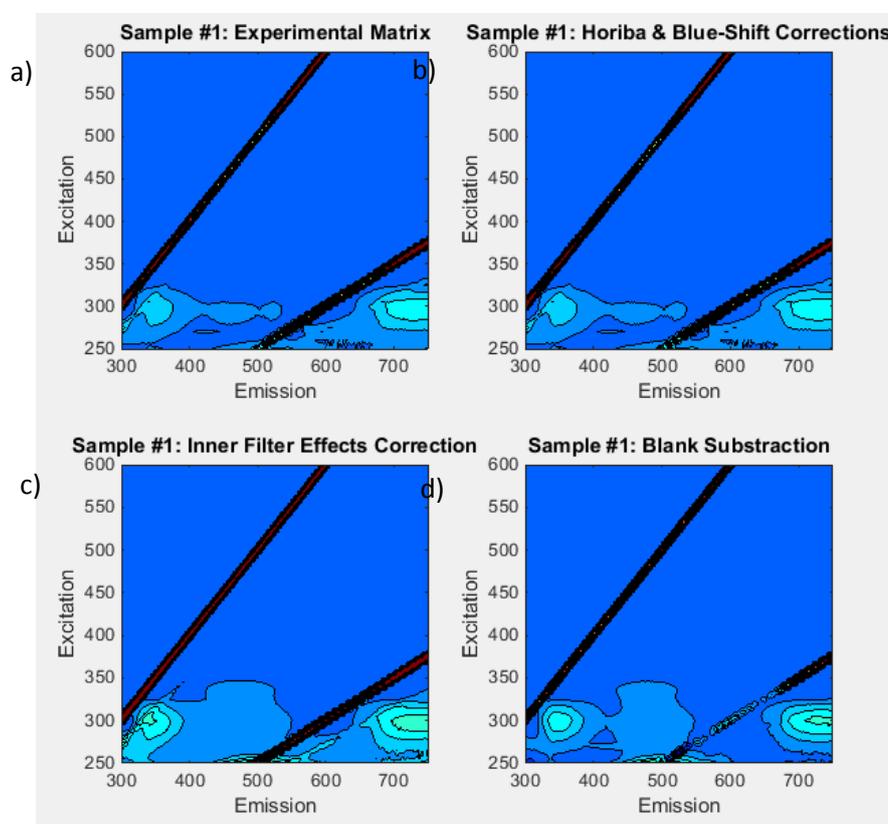


Figura 4.21: (a) MEE sin correcciones. (b) corrección instrumental (c) corrección por absorbancia (d) corrección por blanco.

iv) Normalización

Una vez realizadas las correcciones descritas anteriormente, y luego de verificar que se hayan eliminado las señales que no son propias de la fluorescencia de la muestra, PARAFAC puede ser implementado directamente. Sin embargo, hay que tener en cuenta que cuando el conjunto de datos contiene matrices con valores de intensidad muy dispares entre sí, las muestras con mayor señal tendrán más peso para el ajuste por

cuadrados mínimos que aquellas con baja señal. Para evitar eso, es necesario un paso adicional en el pre-procesado de las muestras. Generalmente cada muestra se normaliza por su área bajo la curva (o intensidad de señal total) de modo de igualar el peso de todas las matrices ante el PARAFAC, permitiendo al modelo centrarse en las variaciones cualitativas entre muestras en lugar de centrarse en la magnitud total de la señal. Cabe destacar que esta normalización se puede “deshacer” después de validar el modelo, multiplicando a cada matriz por el factor de escala que se le aplicó antes de correr el algoritmo.

Fase exploratoria

Una vez finalizado el pre-procesamiento, se realiza una fase exploratoria que tiene la finalidad de establecer el mejor conjunto de datos posible para el modelado y obtener una idea preliminar del número correcto de componentes del modelo. Esta fase incluye la identificación de datos poco representativos o de baja calidad, que se denominan “outliers”, tanto para muestras como para longitudes de onda. Una vez identificados se examinan individualmente para determinar el motivo más probable de su diferencia pudiendo ser eliminados del conjunto de datos. La detección de outliers forma parte del “arte” de PARAFAC y en muchas ocasiones es necesario realizar este paso varias veces ya que resulta crucial para que el ajuste sea satisfactorio.

Para llevar a cabo la identificación de outliers se examinaron en profundidad las matrices de los residuos, los cuales deben distribuirse aleatoriamente o al menos no contener estructuras obvias. Además, se examinó la influencia (leverage) de cada muestra y cada longitud de onda sobre el modelo. El leverage es un número entre 0 y 1 que expresa la desviación con respecto a la media del resto del conjunto de datos, de forma que las muestras/longitudes de onda que son poco diferentes al resto tendrán un leverage cercano a 0 y aquellas muestras/longitudes de onda sustancialmente diferentes al resto tendrán valores cercanos a 1.

Validación del modelo

Existen muchos criterios para evaluar si en un modelo obtenido mediante PARAFAC se ha escogido el número correcto de componentes (factores). A la hora de validar el modelo se han de tener en cuenta el máximo número de criterios de evaluación posibles, puesto que cierto grado de subjetividad es inevitable. En el presente trabajo los criterios tenidos en cuenta a la hora de validar los modelos han sido:

- Falta de ajuste global
- Aleatoriedad de las matrices de los residuos.

- Análisis de la verosimilitud de los perfiles espectrales de excitación y emisión de los componentes.
- Diagnóstico de consistencia del núcleo (Core Consistency).
- Half Split analysis o análisis por mitades

En muchos casos los resultados de un modelo PARAFAC tienen una interpretación química directa, por lo que se debe valorar si los mismos son físicamente razonables con respecto al fenómeno en estudio. En el caso de la espectroscopia de fluorescencia deben tenerse en cuenta las características fundamentales del fenómeno estudiado. Entre otros elementos de juicio, debe considerarse que la energía con la que fluoresce una molécula es generalmente menor que la energía con la que se excita, debido a las pérdidas energéticas que se producen mientras la misma se encuentra en el estado excitado. Por lo tanto, para que los espectros de excitación y emisión obtenidos con el algoritmo sean físicamente aceptables deben presentar las siguientes características generales:]

- Un mínimo de superposición (por lo general <50 nm) entre los espectros de excitación y espectros de emisión.
- Los espectros de excitación puede tener múltiples bandas, pero los espectros emisión correspondientes a un único estado emisor deben exhibir una sola banda.
- Cuando un espectro de excitación tiene dos o más bandas, es decir regiones de absorción correspondientes a estados excitados consecutivos, debe existir absorción, aunque sea mínima, entre estas bandas.
- Los espectros de excitación y emisión generalmente no deben exhibir cambios abruptos en distancias de longitud de onda muy cortas.

Un criterio complementario a la obtención de perfiles físicamente aceptables es el test de evolución de la consistencia del modelo, también llamado consistencia del núcleo. Cuando una secuencia de modelos se ejecuta con un número creciente de factores, el valor de este parámetro tiende a caer bruscamente cuando se emplean más factores que los necesarios. En la práctica la consistencia del núcleo no siempre es un diagnóstico fiable, sobre todo para mezclas complejas con más de 4 o 5 factores.

Por otro lado, en sistemas con varios factores una de las alternativas más eficaces para confirmar si un modelo PARAFAC es apropiado es la validación cruzada basada en el análisis "Half-split". Esta metodología consiste en producir modelos idénticos de submuestras independientes. Esto es, por lo general, sólo posible para conjuntos de datos relativamente grandes. Harshman [71] ha propuesto la validación de modelos utilizando múltiples mitades de pruebas, donde se crean y se comparan varios modelos que luego, de acuerdo a la similitud de los resultados obtenidos con el algoritmo PARAFAC para cada

subconjunto, se determina el número de factores que correlacionan mejor al sistema en estudio.

Capítulo 5. Análisis de espectros de fluorescencia de muestras de orina

5.1 Investigación preliminar con espectros de fluorescencia de muestras de orina

Se realizó una investigación preliminar con la primer orina de la mañana de 15 voluntarios. El grupo se clasificó en personas “sanas” (10), personas afectadas por alguna patología oncológica (4), y un caso particular. En todos los casos al momento de coleccionar la orina las personas se encontraban en ayunas y con un tiempo de retención mayor a 5 horas. Además se les pidió a todos los participantes que prescindan de la ingesta de cualquier medicación durante las 24hs previas al estudio. En esas condiciones fueron coleccionados 20 ml de orina en recipientes estériles, luego conservados a 10°C y analizados dentro de las 24hs.

La población de sujetos sanos incluyó personas de ambos sexos, en igual número, con un rango de edades de 24 a 65 años y diferentes hábitos alimenticios (dos de ellas vegetarianas). Los voluntarios que aportaron las muestras de orina para la población de personas afectadas con patologías oncológicas también fueron de ambos sexos y con edades entre 45 y 69 años. Las patologías que los afectaban incluyeron casos de cáncer de ovario, cáncer de próstata, cáncer de cuello de útero y cáncer renal. Además de estas dos poblaciones bien definidas, se analizó el espectro de la orina de una persona que luego de haber cursado un cáncer de mama y recibir tratamiento oncológico fue diagnosticada como libre de enfermedad.

Todos los espectros de emisión de fluorescencia (EEm) se obtuvieron excitando a 405 nm y manteniendo constantes todos los demás parámetros operativos, como el ancho de la ranura, el voltaje del tubo fotomultiplicador (PMT), la velocidad de barrido, etc. Para facilitar la comparación de las características espectrales de la orina de personas sanas y enfermas, los espectros correspondientes se presentan en una misma figura y normalizados a 1000 u.a. en su máximo.

En la *figura 5.1*, se presentan los resultados obtenidos. En el panel izquierdo se muestra la curva promedio (en negro) y la desviación estándar de los espectros de fluorescencia normalizados, obtenidos de orinas de personas sanas. Dicho espectro presenta un máximo alrededor de 509 nm con un pequeño hombro a 465 nm y dos picos de menor intensidad a 575 y 615 nm. Por otra parte en el panel derecho se comparan los espectros de fluorescencia normalizados correspondientes a orina de personas con distintas patologías oncológicas (CO: Cáncer de ovario, CP: cáncer de próstata, CCU: cáncer

de cuello de útero, CR: cáncer renal), con el espectro promedio de las personas sanas (en negro). Como puede observarse a simple vista, todos los espectros de fluorescencia de las orinas de personas afectadas presentan su máximo alrededor de los 465 nm. Además, en la misma figura, en color amarillo (CMT), se muestra el espectro de la persona que luego de haber recibido tratamiento para el cáncer de mama fue diagnosticada como libre de enfermedad.

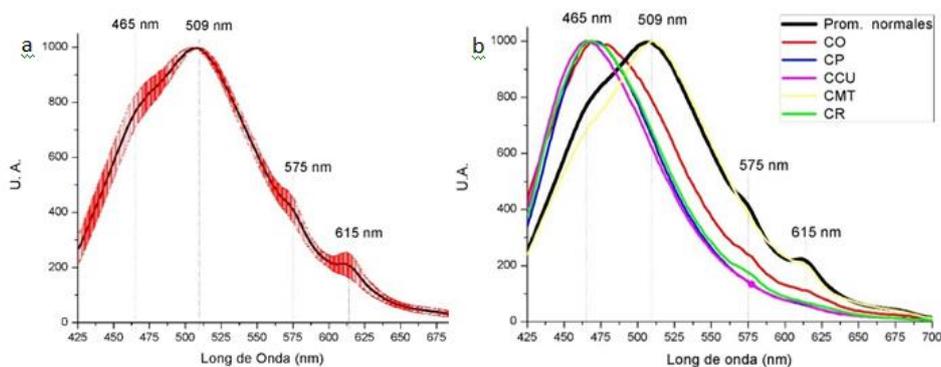


Figura 5.1 (Izq.) Promedio y desviación estándar de los espectros correspondientes a orina de personas sanas. (Der.) Espectros de fluorescencia de orina de las personas con patologías oncológicas (CO: cáncer de ovario, CP: cáncer de próstata, CCU: cáncer de cuello de útero, CR: cáncer renal), cáncer de mama tratado (CMT) y promedio de las sanas (prom. normales).

Los EEm de las muestras de pacientes oncológicos presentan un evidente desplazamiento del máximo de emisión hacia menores longitudes de onda, respecto de los EEm asociados a personas sanas, y prácticamente no se distingue ningún hombro a longitudes de onda menores que la del máximo. Por otra parte, se observa que el EEm de la orina de la paciente con cáncer de mama tratado se corresponde bastante bien con la de los pacientes “sanos”.

Los resultados preliminares obtenidos en esta etapa permiten inferir que la espectroscopia de fluorescencia podría resultar una técnica de “screening” adecuada para detectar enfermedades oncológicas, y que puede fácilmente complementar a otras técnicas. Sin embargo, dada la variabilidad que se observa en los espectros registrados (Figura 5.1), resulta evidente la necesidad de aplicar herramientas estadísticas para el análisis de la información espectroscópica y establecer protocolos precisos tanto para la selección como para la preparación de las muestras con el fin de realizar comparaciones confiables con espectros característicos.

5.2 Ampliación del estudio inicial.

El principal factor limitante que se presentó a la hora de ampliar este estudio fue la necesidad conseguir las muestras de orina apropiadas para la población de “enfermos”. En

el caso ideal, tal como estableció para el estudio preliminar, estas deberían cumplir con ciertas condiciones bien estrictas:

1. Provenir de personas afectadas por alguna patología oncológica, con el diagnóstico confirmado, y no estar bajo ningún tratamiento.
2. No poseer ninguna co-morbilidad (enfermedades coexistentes o adicionales en relación con el diagnóstico oncológico).
3. No haber ingerido ninguna medicación durante las 24 horas previas a la toma de muestra.
4. Colectar la primera orina de la mañana, con un tiempo de retención mayor a 5hs.
5. Estar en ayunas al momento de colectar la muestra

Para cumplir con el ítem 1 se recurrió al servicio de oncología de un hospital público de la ciudad de La Plata, donde luego de una serie de reuniones con el jefe del servicio se acordó que lo más conveniente era invitar a participar en este estudio sólo a los pacientes llamados “de primera consulta”. Es decir a aquellos pacientes que acuden al servicio de oncología por primera vez, derivados de algún otro servicio, para iniciar un tratamiento. De esta manera se convocó a personas que poseían un diagnóstico definido y se descartaron los pacientes que cursaban algún tratamiento.

Respecto de los puntos 2 y 3, y dado que un gran porcentaje de la población incluida en este segmento presentaba también otras co-morbilidades, lo que se resolvió fue que todas las personas que deseaban participar del estudio dejaran sus muestras, independientemente de si cumplían con los dos requisitos previstos. Del mismo modo no se les pidió que alteraran ningún tratamiento en curso, de manera de que no perturbar su medicación habitual. En ese contexto se les solicitó que, junto con el médico, completen una ficha de información detallando con la mayor precisión posible su situación al momento de colectar la muestra, de manera de contar con esa información para complementar el análisis de los espectros.

Los puntos 4 y 5 fueron requisitos durante la primera parte del estudio. Sin embargo, luego de transcurridos varios meses la cantidad de muestras recibidas en el laboratorio fue muy escasa. En la mayoría de los casos, las personas al momento de ir al hospital a su primer consulta no estaban en ayunas, o bien no contaban con un tiempo de retención suficiente, lo que impedía su participación en el estudio en ese momento. La necesidad de volver al hospital otro día para llevar la muestra resultaba en una gran incomodidad para el paciente. Se resolvió entonces cambiar el criterio para la selección de muestras y admitir que los voluntarios recolecten la muestra el mismo día que se los invitaba a participar, luego de la consulta y en la situación que se encontraran, indicando la cantidad de horas de ayuno y de retención. De esta manera se logró colectar una cantidad importante de muestras, que luego de procesar la información espectral, fueron

clasificadas según las condiciones de colección que figuraban en la ficha adjunta. Además, para evaluar la influencia del incumplimiento de los requisitos de estos dos puntos, se realizó un análisis de la variabilidad intra-individuo en diferentes condiciones de ayuno y retención, que se presenta en la sección 5.4.

En esta segunda etapa se obtuvieron y midieron más de 120 muestras de orina de pacientes oncológicos y 60 de personas “sanas”. Para el análisis que se presenta a continuación no se tuvieron en cuenta las muestras provenientes de personas que cursaban tratamientos por hipertensión al momento de coleccionar la muestra ni aquellas diagnosticadas con diabetes. Tampoco las que cursaban más de dos co-morbilidades o tomaban mucha medicación. Por otra parte, avalados por los resultados que se describen en la sección 5.4 (que muestran que la variabilidad intra-individuo para diferentes condiciones de ayuno y retención, es menor a la variabilidad entre-individuos), se incluyeron las muestras que no fueron recolectadas de la primer orina de la mañana. De esta manera se analizaron 81 muestras en total: 37 de personas sanas y 44 de personas con patologías oncológicas. En la figura 5.2 puede observarse la totalidad de los espectros, normalizados a 1000 u.a. en su máximo. En color verde se muestran los correspondientes a personas sanas y en color rojo los de las orinas de personas con alguna patología oncológica.

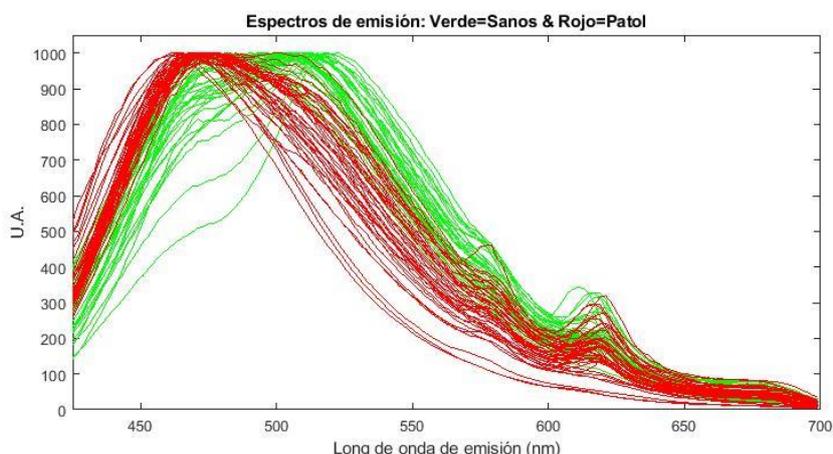


Figura 5.2 EEm normalizados al pico máximo de cada uno, correspondientes a 37 muestras de personas sanas (en verde) y 44 de personas con patologías oncológicas (en rojo).

Si bien las muestras presentan una gran variabilidad, puede observarse que los resultados concuerdan con los de la investigación preliminar. Los espectros correspondientes a personas que cursan procesos oncológicos tienden a tener su máximo desplazado hacia menores longitudes de onda, respecto de los correspondientes a personas sanas.

Para poder observar con mayor detalle las características globales de ambas poblaciones, en la figura 5.3 se presenta en verde el promedio de los espectros de fluorescencia normalizados de las 37 orinas de personas sanas, y en rojo el espectro promedio de las 44 muestras de personas con patologías oncológicas.

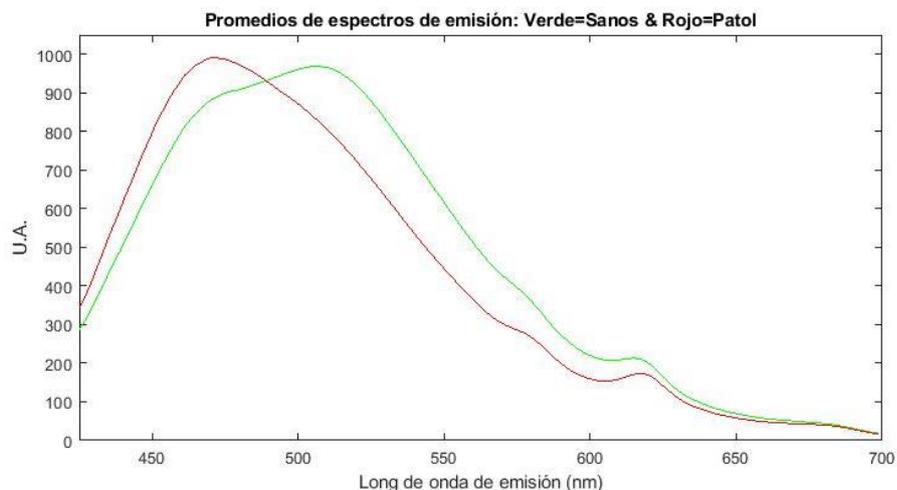


Figura 5.3 Promedio de los EEm de 37 muestras de orinas de personas sanas (en verde), y de las 44 muestras de personas con patologías oncológicas (en rojo).

El máximo de intensidad de fluorescencia para las muestras de personas sanas ronda los 507 nm, con un pequeño hombro alrededor de los 470 nm y dos picos de menor intensidad a 575 y 617 nm. En contraste, los espectros de las personas afectadas presentan un evidente desplazamiento del máximo de emisión hacia menores longitudes de onda, mostrando en promedio un máximo alrededor de los 470 nm, y prácticamente no se distingue ningún hombro a longitudes de onda menores que la del máximo. La pendiente desde el máximo hacia mayores longitudes de onda parece ser menos pronunciada que la del promedio de las curvas normales, mientras que los picos secundarios a mayores longitudes de onda no parecen presentar diferencias significativas entre ambos espectros promedio.

Si bien se podría pensar que la diferencia entre los promedios de ambas poblaciones a cada longitud de onda es el parámetro de comparación ideal, es importante destacar que eso solo no es suficiente ya que la variabilidad también debe ser considerada. Por ejemplo, en los casos en los cuales los valores registrados de intensidad de fluorescencia de cada población presentan una gran variabilidad, puede suceder que dicha variabilidad resulte comparable a la diferencia entre las medias de ambas poblaciones. Para ilustrar este punto en la figura 5.4 se presentan las curvas promedio de la figura 5.3 junto con sus desviaciones estándar. Las desviaciones estándar, que indican qué tan dispersos están los datos con respecto del promedio a cada longitud de onda particular, se representaron con barras verticales del mismo color en cada una de las curvas. Tal como puede observarse, si bien los promedios de ambas poblaciones toman

valores diferentes, en la zona entre 465 y 485 nm (recuadro azul) las barras que representan las desviaciones estándar quedan superpuestas. Por lo tanto, la diferencia entre medias por sí sola no tiene sentido en cuanto a la capacidad de diferenciar entre espectros de personas sanas y afectadas. En este contexto, es sumamente importante tener en cuenta el compromiso que existe entre la separación de los valores medios y variabilidad dentro de cada población. Este punto introducido aquí será tenido en cuenta y discutido a lo largo de todo el Capítulo.

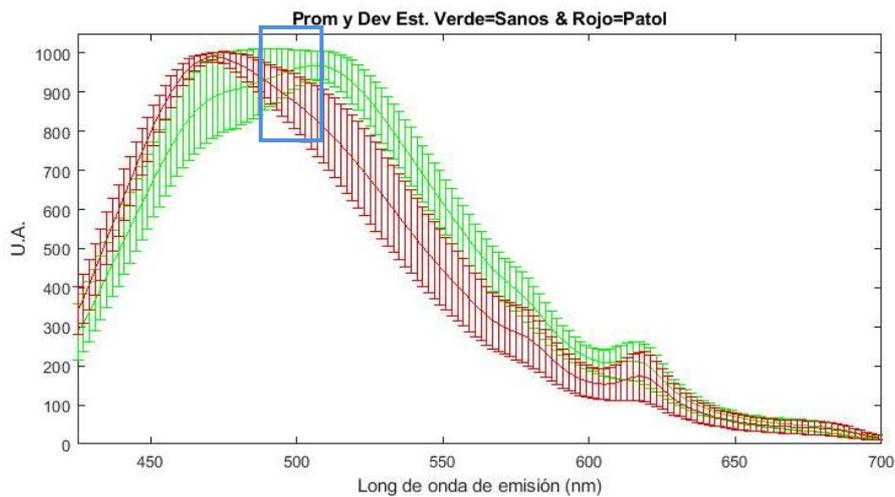


Figura 5.4 Promedio y desviación estándar de los EEm de 37 muestras de orinas de personas sanas (en verde), y de las 44 muestras de personas con patologías oncológicas (en rojo).

Si bien la comparación de la información espectral mediante su representación gráfica puede aportar cierta información, es una metodología de análisis cualitativa y puede resultar muy subjetiva. Cada espectro de emisión de fluorescencia puede ser considerado como un vector de 138 dimensiones, donde cada una de ellas corresponde al valor de la intensidad de emisión de fluorescencia de esa muestra a la i -ésima longitud de onda de emisión λ_i , con $425 \leq \lambda_i \leq 699$. Desde este punto de vista se puede pensar, entonces, en realizar algún tipo de comparación numérica entre poblaciones. Si bien esto es más adecuado que la comparación gráfica, se hace evidente que analizar y comparar un número elevado de muestras de estas dimensiones resulta complicado. En consecuencia, en una etapa inicial del estudio se buscaron índices de interés, o valores representativos de los espectros para comparar y tratar de establecer si existen diferencias entre la fluorescencia de las muestras de orina de ambas poblaciones. De aquí en adelante con la palabra “índice” se hará referencia a algún parámetro extraído de cada uno de los espectros capaz de discriminar una población de otra y que, posteriormente, permita desarrollar un test. En otras palabras, se intentó determinar en primer lugar si la fluorescencia de la orina es capaz de reflejar la existencia de dos poblaciones, para luego, en un paso posterior determinar cuál índice o valor es el que mejor discrimina a los individuos de cada población.

Pueden encontrarse en bibliografía varios trabajos que han estudiado la posibilidad de diagnosticar enfermedades a partir del análisis de la fluorescencia de algún fluido biológico como sangre, orina, suero, plasma, etc. [72-76] En general, en la mayoría de ellos se utilizaron como índices de interés cocientes intensidades de fluorescencia a longitudes de onda de emisión fijas para todos los espectros. El criterio de selección de dichas longitudes de onda es propio de cada análisis y se basa en alguna característica particular de la zona espectral considerada. Esto es muy sencillo de realizar y tiene la ventaja de que, al ser una medida relativa, es independiente del tipo de normalización empleado y no es necesario tener en cuenta la intensidad total de fluorescencia (ni calibrar el equipo). Empleando esta misma estrategia se generaron varios índices de interés, que se presentan a continuación. Sin embargo, y a diferencia de la mayoría de los trabajos publicados al momento, en el presente estudio se realizó un análisis estadístico detallado de cada índice para determinar su capacidad de evidenciar la existencia de dos poblaciones, y posteriormente establecer cuál de los índices presentaba el mejor poder de discriminación entre ellas. Además se desarrollaron e implementaron algoritmos para seleccionar en forma automática las longitudes de onda que pudieran funcionar como buenos índices, tal como se verá a continuación.

Tomando en consideración la estrategia usualmente empleada en las publicaciones más recientes se seleccionaron las longitudes de onda que se listan en la tabla 5.1 para realizar los cocientes. En la misma tabla se detalla el motivo por el cual cada longitud de onda fue seleccionada.

λ_{emis}	Característica de la zona
471 nm	Zona del pico máximo de fluorescencia de las muestras patológicas
507 nm	Zona del pico máximo de fluorescencia de las muestras sanas
540 nm	Pendiente
575 nm	Pico suave
617 nm	Pico suave
675 nm	Zona de menor variabilidad entre muestras (o de mayor similitud)

Tabla 5.1 Longitudes de onda de emisión seleccionadas, junto con sus características

Los índices se calcularon entonces como cocientes entre las intensidades de fluorescencia para las longitudes de onda de la tabla 5.1, de modo que el índice $I_{XXX/YYY}$ es aquel índice generado al tomar el cociente de la emisión a XXX sobre la emisión a YYY (por ejemplo $I_{471/575}$). Esta misma notación se mantiene a lo largo de todo el Capítulo.

Los valores de los índices obtenidos para cada muestra se muestran en las figura 5.5. El índice en estudio se identifica en el título de cada recuadro. En todos los casos, para

facilitar la visualización, se graficaron superpuestos los valores obtenidos para los índices de personas sanas (en verde), y los de las personas afectadas (en rojo), junto con sus promedios (líneas sólidas).

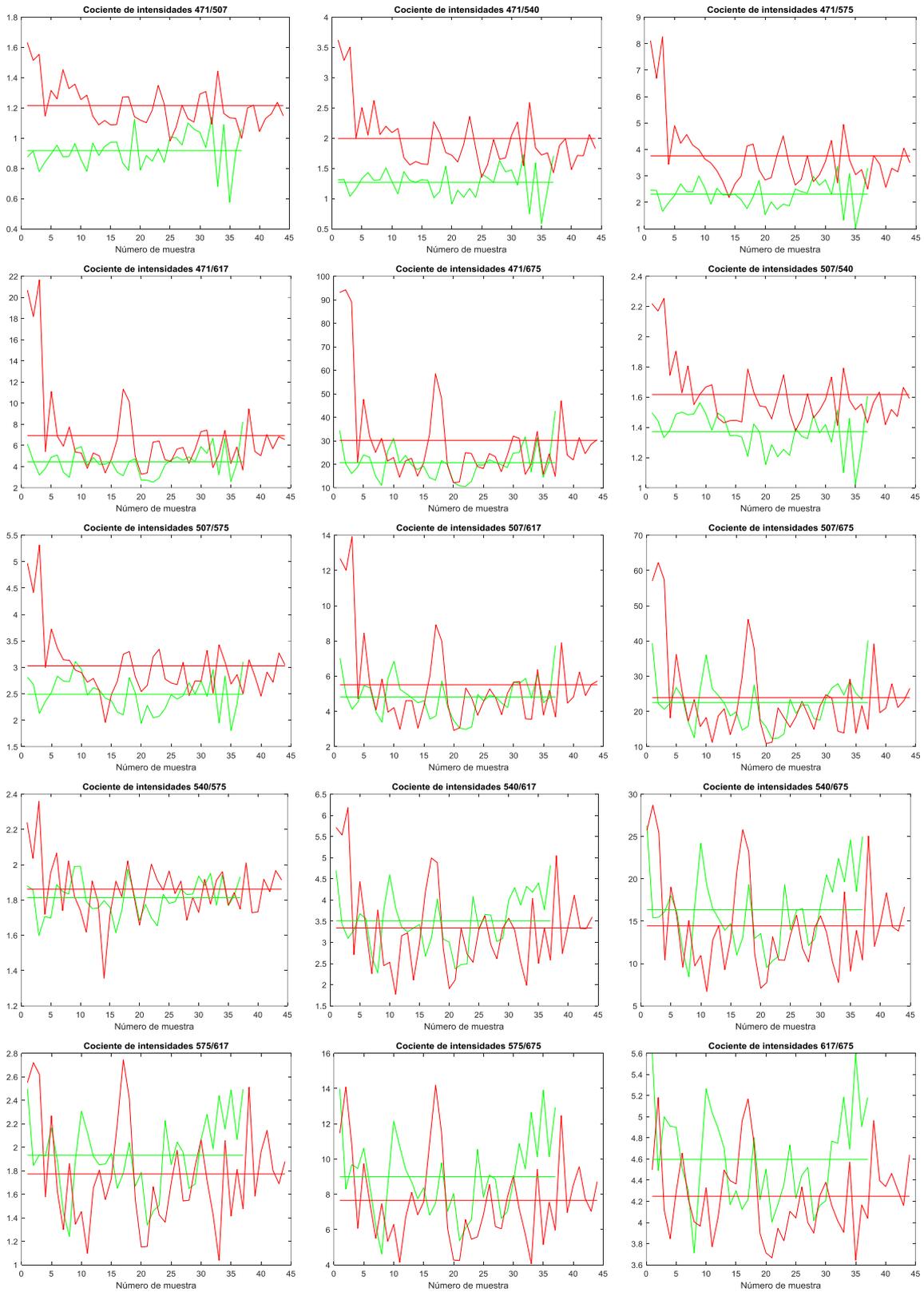


Figura 5.5 Valores de los índices de la tabla 5.1 para cada muestra y promedio de ambas poblaciones.

Tal como puede observarse algunos índices (como $I_{617/675}$, y $I_{507/675}$) no son capaces de distinguir la existencia de dos poblaciones, mientras que otros (como $I_{471/507}$, y $I_{471/575}$) parecen evidenciar la presencia de dos poblaciones diferentes. Aquí también se observa la importancia de considerar el compromiso entre la separación de los valores medios y variabilidad dentro de cada población.

Teniendo en claro que cualquier método diagnóstico debe estar basado en un completo análisis estadístico, se planteó la necesidad de evaluar si la diferencia en la distribución de cada uno de esos índices para personas sanas y personas con patologías oncológicas tiene significación estadística. Posteriormente se realizó una valoración de la calidad que tiene cada uno de esos índices para discriminar las poblaciones.

5.3 Análisis estadístico de los datos

Dada la importancia de realizar un riguroso análisis estadístico de los índices obtenidos, como parte del presente trabajo de Tesis se desarrolló un software específico en Matlab. Con el objetivo final de comparar la potencia estadística de los índices generados, dicho software realiza múltiples cálculos para obtener parámetros característicos de cada índice entre los que se destacan los estadísticos de Fisher (F), de Student (t) y de Welch, descritos en el Capítulo 4, y sus correspondientes grados de libertad. Además, para facilitar el análisis de normalidad genera las curvas Q-Q para cada población [77] y presenta los histogramas de cada población superpuestos a las distribuciones normales estimadas a partir de los respectivos parámetros de media y dispersión. El número de categorías a usar para generar el histograma puede ser fijado en forma manual, o bien calculado automáticamente por el software como el número entero más próximo la raíz cuadrada del número de muestras. El software fue diseñado además para generar la curva ROC correspondiente al índice en estudio, pudiendo seleccionar manualmente la cantidad de corrimientos de la línea umbral. A partir de la curva ROC el software estima el área debajo de la misma (AUC) y también calcula las matrices de contingencia para ese test. Haciendo uso de esta rutina se analizaron los índices de la sección anterior y se obtuvieron los valores de los parámetros característicos que se muestran en la tabla 5.2. En la primera columna se indica el índice en cuestión, cuyo nombre está compuesto por las dos longitudes de onda de emisión con las que se calculó el cociente separadas por una barra. La segunda columna lista los valores de la relación señal/ruido (S/N, por sus siglas en inglés) para cada índice, mientras que en la tercera se muestran los valores del parámetro de Fisher, con sus respectivos grados de libertad entre

paréntesis. En las columnas 4 y 5 se muestran los valores del Test de Student y de Welch, mientras que la columna 6 lista los valores de área bajo la curva (AUC) ROC.

1	2	3	4	5	6	7
Índices	S/N	F-Fisher	t-Student (<i>f</i>)	t-Welch (<i>f'</i>)	AUC	
I _{471/507}	2,248	1,207(43,36)	10,078 (79)	10,173 (79)	0,954	
I _{471/540}	1,783	3,627(43,36)	7,996 (79)	8,391 (64)	0,948	
I _{471/575}	1,453	5,071(43,36)	6,512 (79)	6,889 (59)	0,914	
I _{471/617}	0,790	8,502(43,36)	3,543 (79)	3,783 (53)	0,756	✗
I _{471/675}	0,627	6,653(43,36)	2,809 (79)	2,987 (55)	0,670	✗
I _{507/540}	1,414	2,047(43,36)	6,337 (79)	6,532 (74)	0,861	
I _{507/575}	1,064	2,922(43,36)	4,769 (79)	4,974 (68)	0,791	✗
I _{507/617}	0,356	4,224(43,36)	1,597 (79)	1,683 (62)	0,530	✗
I _{507/675}	0,146	2,763(43,36)	0,654 (79)	0,681 (69)	0,449	✗
I _{540/575}	0,352	2,257(43,36)	1,580 (79)	1,634 (73)	0,581	✗
I _{540/617}	0,195	2,078(43,36)	-0,872 (79)	-0,899 (74)	0,585	✗
I _{540/675}	0,380	1,311(43,36)	-1,705 (79)	-1,727 (79)	0,633	✗
I _{575/617}	0,413	1,604(43,36)	-1,852 (79)	-1,891 (77)	0,632	✗
I _{575/675}	0,562	1,032(43,36)	-2,521 (79)	-2,528 (79)	0,663	✗
I _{617/675}	0,857	1,647(43,36)	-3,840 (79)	-3,763 (71)	0,716	✗

Tabla 5.2 Valores estimados de los estadísticos.

La estimación de la relación señal/ruido de cada índice, se realizó a partir del cálculo del cociente entre la diferencia entre las medias muestrales de cada población (enfermos - sanos) y la varianza combinada. Cabe aclarar que se calculó de esta manera asumiendo la hipótesis de que existen dos poblaciones con medias diferentes, pero con dispersiones similares. Un valor del cociente S/N menor que la unidad, significa que ese índice no refleja ninguna información respecto a la existencia de dos poblaciones por sobre el ruido. Los valores obtenidos se muestran en la segunda columna de la tabla 5.2. Como es de esperar, los índices que presentan valores de S/N bajos, son los mismos que presentan ambas curvas solapadas en la figura 5.5. En consecuencia, se descartaron los índices que se indican con una ✗ en la última columna de la tabla 5.2, por presentar valores del cociente $S/N \leq 1,1$.

Si bien los índices no descartados en el paso anterior indicaban que podía existir una diferencia entre las medias muestrales de enfermos y sanos, el siguiente paso fue

decidir si la diferencia obtenida para cada uno era estadísticamente significativa. En otras palabras, se testeó, para cada índice, si las muestras de sanos y enfermos correspondían a distribuciones de probabilidad con medias poblacionales distintas. En caso contrario se asumió que la diferencia entre medias podía deberse a variaciones estadísticas al azar, y en ese caso se descartó el índice en cuestión. Para hacer esta discriminación se utilizó el test t-Student o el test de Welch (en el caso de tener varianzas homogéneas o heterogéneas respectivamente), explicados en la sección 4.3. Dado que ambos test son válidos sólo si las observaciones de cada grupo presentan una distribución aproximadamente normal, también se analizó ese punto en particular. Si bien dada la cantidad de muestras de cada grupo se podría asumir la hipótesis de normalidad, se construyeron los histogramas de cada población para verificarla. En las figuras 5.6 y 5.7 se muestran los histogramas obtenidos de las muestras de pacientes sanos (verde), de pacientes enfermos (rojo) y de ambas poblaciones juntas (en negro) para cada índice. Superpuestos a éstos, el software desarrollado presenta una campana de Gauss con los valores de μ y σ estimados a partir de cada una de las subpoblaciones, siguiendo la misma convención de colores. De esta manera se realizó un análisis de “normalidad” por comparación de cada histograma su curva correspondiente.

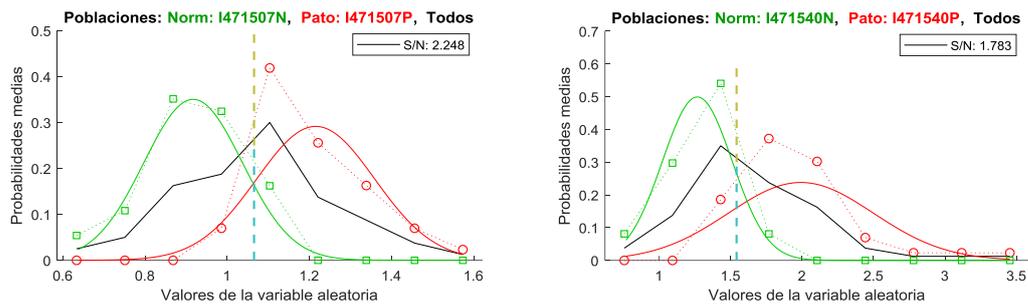


Figura 5.6 Histogramas correspondientes a los índices 471/507 (Izq.) y 471/540 (Der.).

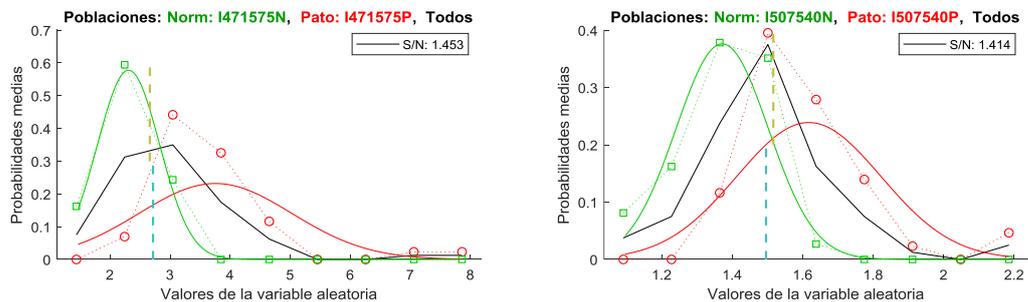


Figura 5.7 Histogramas correspondientes a los índices 471/575 (Izq.) y 507/540 (Der.).

Verificada la normalidad de las distribuciones se realizó la prueba F de Fisher para analizar la homogeneidad de las varianzas, y posteriormente decidir que test usar para comparar las medias (los valores F obtenidos se muestran en la columna 3 de la tabla 5.2).

En los casos en los cuales la prueba de Fisher no indicó diferencias significativas entre las varianzas con un $\alpha=0,01$ ($F_{obtenido} < F^{0,01}_{n-1,m-1}$), entonces se asumió la homogeneidad de las varianzas y se usó el test de Student (t-test) convencional. En cambio, para los casos en que $F_{obtenido} > F^{0,01}_{n-1,m-1}$, se asumió la existencia de diferencias significativas entre las varianzas, por lo que se usó el test de Welch.

A modo de ejemplo a continuación se describen algunos resultados particulares. Por ejemplo, el estadístico F para las varianzas del índice $I_{471/507}$ presentó un valor $F_{43,36}=1,207 < F_{crítico} (F^{0,01}_{43,36}=2,16)$, por lo que se asumió la homogeneidad de las varianzas, y se evaluó la diferencia entre las medias poblacionales con el test de Student. Dado que el valor t obtenido fue 10,078 (Tabla 5.2, columna 4), mucho mayor al $t_{crítico} (t^{0,01}_{79}=2,379)$, pudo afirmarse que existe una diferencia significativa entre las medias de este índice para pacientes sanos y enfermos ($p < 0,0005$). Por otra parte, a diferencia del caso anterior, el valor de estadístico F obtenido para las varianzas del índice $I_{471/540}$ ($F_{43,36}=3,627$) fue mayor al $F_{crítico}$, por lo que en este caso no pudo asumirse la homogeneidad de las varianzas y se recurrió al test de Welch para la comparación de las medias. El valor del estadístico de $t_{Welch}=8,391$ obtenido (tabla 5.2 columna 5), fue mucho mayor al t de Student de 64 grados de libertad ($p < 0,01$), por lo que se concluyó que con este índice también existen diferencias significativas entre las medias para sanos y enfermos.

De todos los resultados obtenidos se concluyó que los índices $I_{471/507}$, $I_{471/540}$, $I_{471/575}$ e $I_{507/540}$ extraídos de la fluorescencia de las muestras de orina tienen la capacidad de discriminar entre personas sanas y pacientes afectados por procesos oncológicos, ya que corresponden a distribuciones de probabilidad con medias poblacionales distintas. El siguiente paso fue emplear el software desarrollado para evaluar la eficacia de cada índice para desarrollar un test diagnóstico capaz de lograr esa discriminación y evaluar el punto de corte óptimo para la clasificación. Para ello se realizaron las correspondientes curvas ROC, que se muestran en las figuras 5.8 y 5.9.

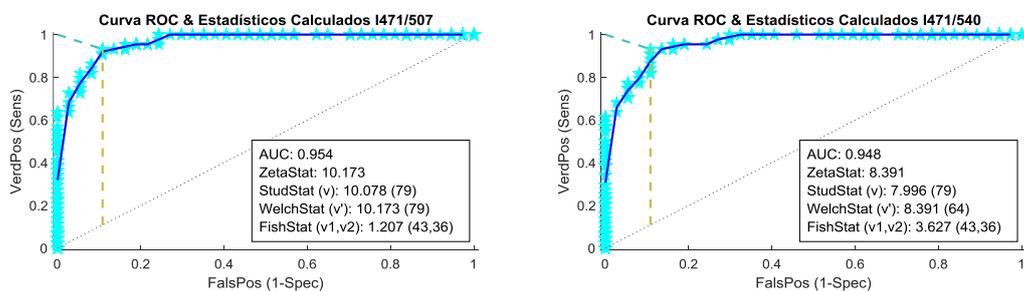


Figura 5.8 Curvas ROC correspondientes a los índices $I_{471/507}$ (Izq.) y $I_{471/540}$ (Der.).

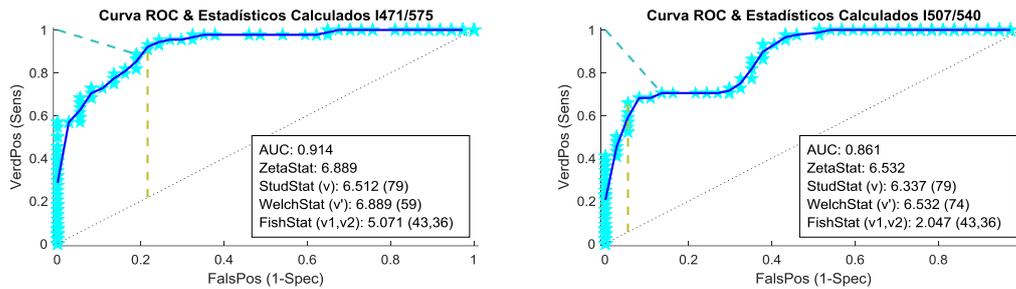


Figura 5.9 Curvas ROC correspondientes a los índices $I_{471/575}$ (Izq.) y $I_{507/540}$ (Der.).

En cada figura se muestra la curva ROC "cruda" en estrellas celestes, y la curva ROC suavizada por intervalos de especificidad, en trazos azules. La diferencia entre estas curvas radica en que cuando más de un punto tiene la misma especificidad, la curva ROC "cruda" toma el máximo valor de sensibilidad obtenido, mientras que la suavizada es aquella que considera el valor medio de las sensibilidades obtenidas para dicha especificidad. Para ejemplificar este concepto en la figura 5.10 se muestran las curvas ROC generadas según cada criterio, donde se destacan con elipses negras los conjuntos de puntos de igual especificidad que se ajustan según la línea azul en cada caso. Cabe destacar que a lo largo del presente trabajo se calculó el AUC a partir de la curva ROC suavizada por ser un criterio más exigente que el de la ROC sin suavizar. En la caja de texto de la esquina inferior derecha de las figuras 5.8 y 5.9 se muestran el valor del AUC (calculado a partir de la curva ROC suavizada por intervalos de especificidad), y los valores calculados para los estimadores estadísticos de interés.

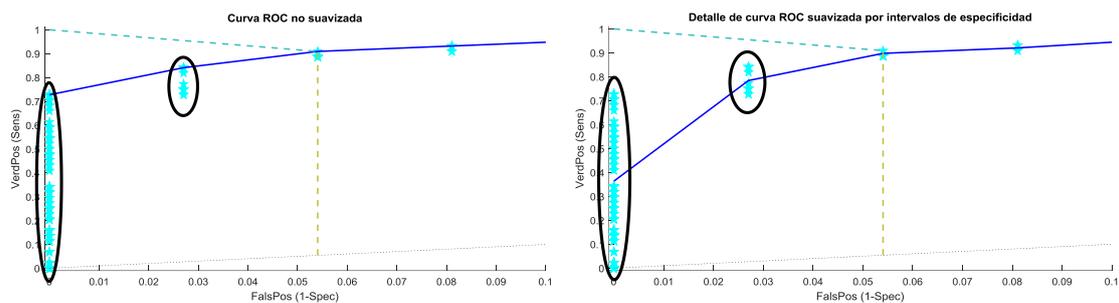


Figura 5.10 Curvas ROC no suavizada (Izq.), y suavizada por intervalos de especificidad (Der.) correspondientes al mismo conjunto de datos.

Tal como se explicó en la sección 4.3, el área total bajo la curva ROC (AUC) permite ordenar en forma combinada la sensibilidad y la especificidad del test, y representa un índice único para estimar el rendimiento de una prueba. En líneas generales, cuanto mayor es el AUC, mejor será el rendimiento general de la prueba de diagnóstico para clasificar correctamente los sujetos enfermos y no enfermos. La comparación de los valores de AUC

para los índices presentados en las figuras 5.8 y 5.9 muestra que, en principio, el índice que reflejaría el mejor rendimiento es el denominado $I_{471/507}$ (AUC=0,954), es decir el cociente de las intensidades a las longitudes de onda en las que mayormente se presentan los máximos de fluorescencia de las muestras de orina de pacientes enfermos y sanos respectivamente. En orden decreciente, en segundo lugar se encontraría el índice $I_{471/540}$ (AUC=0,948), luego el $I_{471/575}$ (AUC=0,914) y, por último, el $I_{507/540}$ (AUC=0,861). Es importante recordar que el AUC puede tomar valores entre 0,5 (test sin valor diagnóstico), y 1 (test de valor diagnóstico perfecto). Por lo tanto los valores de AUC obtenidos a partir de estos cuatro índices muestran que los test analizados pueden ser clasificados como “buenos” (AUC entre 0,75 y 0,90) y “muy buenos” (AUC entre 0,90 y 0,97).

Nuevos índices de interés

A pesar de que la estrategia habitualmente empleada en la bibliografía publicada sobre el tema ha permitido generar una serie de índices de interés, es muy importante tener en cuenta que cada espectro completo contiene una gran cantidad de información que, al tomar sólo los valores de emisión de fluorescencia a una o dos longitudes de onda solamente, no es tenida en cuenta. Es por ello que uno de los objetivos particulares del presente estudio se ha centrado en la búsqueda de algún nuevo índice o método que pueda aprovechar una mayor cantidad de la información contenida en los espectros medidos para mejorar la calidad del test.

En este contexto, consideramos que una contribución fundamental de este trabajo radica en evaluar diferentes estrategias de análisis que, incluyendo la mayor cantidad de información que pueda ser relevante del conjunto de datos registrados, permitan desarrollar tests con capacidad diagnóstica. En este sentido, el presente trabajo se basa en evaluar diferentes formas de transformar matemáticamente los datos de los espectros de fluorescencia de orina para hallar algún índice que sea capaz de discriminar entre individuos de ambas poblaciones.

A lo largo del presente trabajo se han testeado más de cien índices diferentes. A continuación se describen algunos de ellos, que fueron seleccionados por ser los que mejor servirían para discriminar ambas poblaciones. En particular, en esta sección se analizan los que surgen a partir de valores únicos de cada espectro (con diferentes criterios de normalización, por ejemplo), y operaciones matemáticas sencillas (tales como resta, producto y cociente).

En una primera etapa los índices se calcularon a partir de los espectros normalizados al área bajo la curva (área =1), y de ahí que se los denominó con el prefijo “Area” para su descripción. De todos los índices probados en esta primera etapa, los cuatro que mostraron la mayor capacidad para separar ambas poblaciones son:

- $I_{Area471}$: el valor de la intensidad de emisión en 471nm, correspondiente a la zona del pico máximo de fluorescencia de las muestras patológicas;
- $I_{Area507}$: el valor de la intensidad de emisión en 507nm, correspondiente a la zona del pico máximo de fluorescencia de las muestras sanas;
- $I_{Area471-507}$: la diferencia entre los dos índices recién mencionados, y
- I_{PosMax} : el valor de la longitud de onda en la que se presenta el máximo de cada espectro.

En la figura 5.10 se muestran los resultados obtenidos empleando estos cuatro índices.

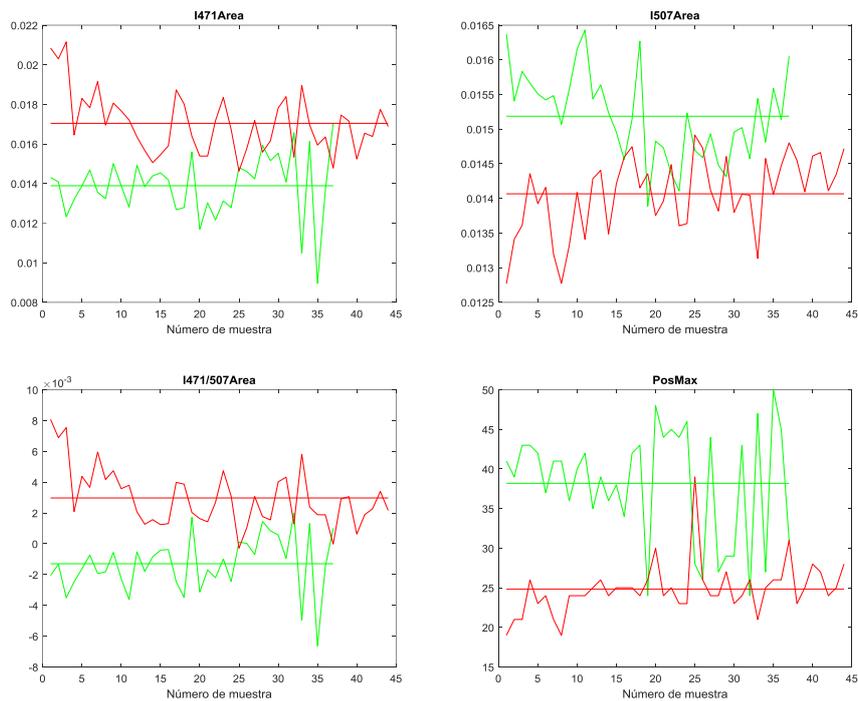


Figura 5.10 Valores de los índices. $I_{471Area}$ $I_{507Area}$ $I_{471/507}$ e I_{PosMax} . y promedio de cada población. El índice en estudio se indica en el título de cada recuadro.

En la tabla 5.3 se listan los valores obtenidos a partir del análisis estadístico para este grupo de índices, y las figuras 5.11 a 5.14 muestran los histogramas y curvas ROC correspondientes.

Índices	S/N	F-Fisher	t-Student	t-Welch	(p)	AUC
$I_{Area471}$	2,016	1,279 (36,43)	9,039	8,954	76	0,928
$I_{Area507}$	1,929	1,595 (36,43)	8,649	8,486	72	0,904
$I_{Area471-507}$	2,3	1,143 (36,43)	10,309	10,262	77	0,952
I_{PosMax}	2,461	5,913(36,43)	11,033	10,343	48	0,920

Tabla 5.3 Valores estimados de los estadísticos

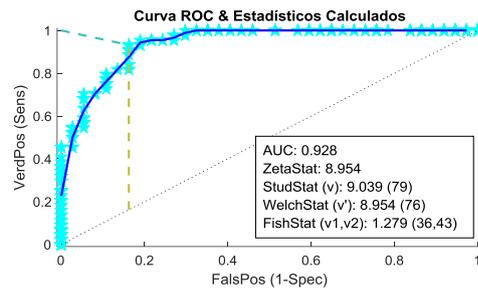
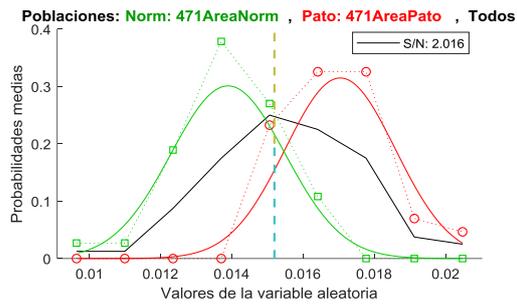


Figura 5.11 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice $I_{Area471}$.

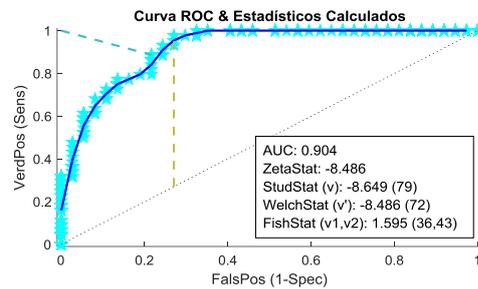
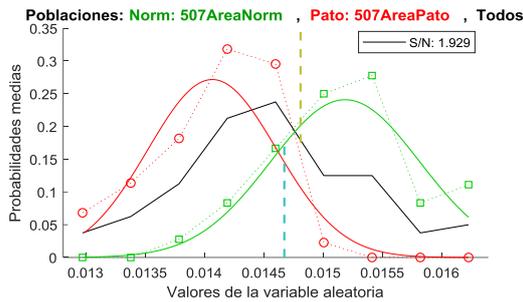


Figura 5.12 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice $I_{Area507}$.

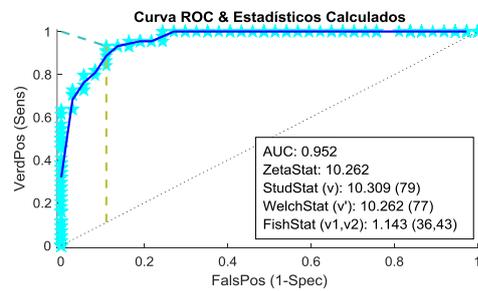
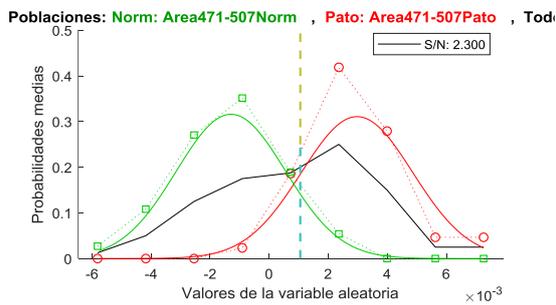


Figura 5.13 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice $I_{Area471-507}$.

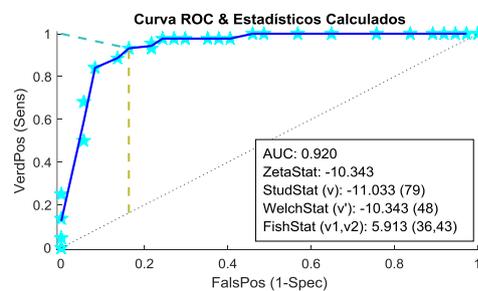
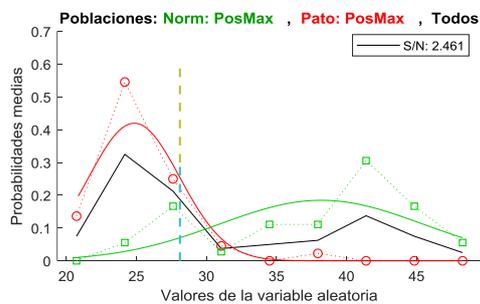


Figura 5.14 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice $I_{PosMáx}$.

Los resultados muestran claramente que, dependiendo del nivel de confianza seleccionado, algunos de estos índices evidencian la existencia de 2 poblaciones (enfermos y sanos). Esta sencilla evaluación se ha realizado a partir de los valores de emisión de fluorescencia obtenidos a una o dos longitudes de onda y este tipo de estrategia es lo único que ha sido publicado en bibliografía hasta el momento. Cabe destacar además, que si bien en las publicaciones existentes se han calculado las correspondientes curvas ROC para los diferentes índices, éstos últimos no han sido analizados en detalle mediante ningún test estadístico.

5.4 Análisis por componentes principales (PCA)

Tal como se explicó en la sección 4.8, el análisis de componentes principales (PCA) es una técnica que se utiliza para reducir la dimensionalidad de un conjunto de datos. El objetivo del PCA es transformar una serie de variables posiblemente correlacionadas entre sí en un número menor de variables no correlacionadas denominadas componentes principales (PCs). A continuación se describe cómo se aplicó esta técnica para lograr la reducción de la dimensionalidad de los datos, de manera de contribuir a la visualización de tendencias y facilitar la discriminación entre poblaciones. Con ese objetivo se generó una rutina específica en Matlab cuyo primer paso es justamente una reducción de la dimensionalidad mediante PCA. De esta manera se obtuvo una proyección del espacio de n dimensiones ($n=138$ en este caso), en un espacio bidimensional, que conserva el 93,5% de la variabilidad total. Es decir, con sólo dos dimensiones, dos índices por muestra, se logró capturar más del 90% de las características espectrales que resultan de interés. Esta reducción al plano PC1/PC2 se hizo considerando todos los espectros: los correspondientes a ambos conjuntos de pacientes ("sanos" y "patológicos").

El siguiente paso consistió en calcular las "elipses de confianza" o "elipses de error de covarianza". Para ello se asumió que las diferentes poblaciones (enfermos, sanos y total) poseían una distribución normal en dos dimensiones. Es decir, se consideró que cada población se podía representar aproximadamente como una distribución normal en dos dimensiones (correspondientes a los ejes PC1 y PC2), con sus correspondientes σ_i , σ_j , μ_i y μ_j . En este contexto, las elipses de confianza que se calcularon representan un isocontorno de la distribución poblacional que se asumió como gaussiana bidimensional y define la región que contiene un cierto porcentaje $(1-\alpha)$ de todas las muestras que se pueden extraer de la población considerada. Es decir, las elipses calculadas definen una región con un determinado nivel de confianza pre fijado. La gran utilidad de estas elipses radica en que permitieron visualizar un intervalo de confianza en dos dimensiones. La figura 5.15 muestra, a modo de ejemplo, tres elipses con niveles de confianza $(1-\alpha = 90, 95$ y $99)$ para un conjunto de datos distribuidos normalmente en 2D.

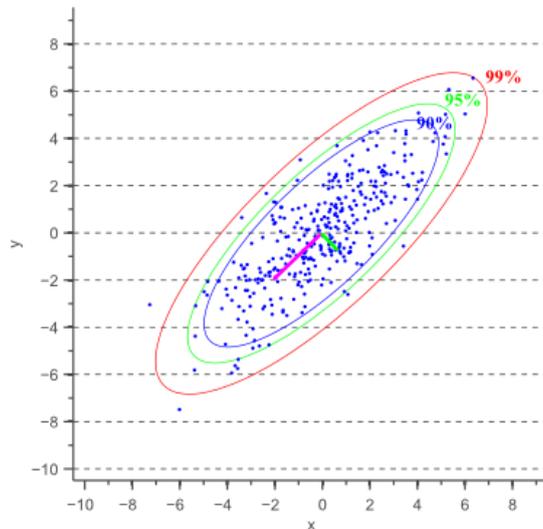


Figura 5.15 ($1-\alpha$) es valor porcentual que aparece en cada elipse y representa el nivel de confianza con el que se dibujan las elipses correspondientes a cada población. Azul: $1-\alpha=90\%$, verde $1-\alpha=95\%$, rojo $1-\alpha=99\%$.

Para analizar las diferentes elipses, en primer lugar se calculó la matriz covarianza del conjunto de datos, considerando sólo una población, que contempla tanto muestras de personas sanas como enfermas. Los autovectores de dicha matriz proporcionan las direcciones de los semiejes de la elipse de la población total en el espacio bidimensional PC1/PC2. Los autovalores de esta matriz dan cuenta de la dispersión de los puntos de la población respecto del eje principal. Cuanto mayor sean los autovalores, mayor será la dispersión característica de la población respecto de su eje principal. En otras palabras, los autovalores representan la magnitud de la varianza de los datos en la dirección de los autovectores.

El ángulo de inclinación de la elipse está determinado por la covarianza de los datos. Si la covarianza es cero (caso en el que los datos no están correlacionados), la elipse estará alineada con los ejes. En ese caso los autovalores son iguales a las varianzas de la matriz de covarianza y los autovectores son iguales a la definición del eje x y del eje y. En el caso de datos correlacionados, los autovectores representan las direcciones de mayor dispersión de los datos, mientras que los autovalores definen qué tan grande es realmente esta extensión. Como es de esperar, las magnitudes de los ejes de la elipse dependen de la varianza de los datos.

La rutina desarrollada permite graficar todos los puntos en el espacio PC1/PC2, con la elipse calculada superpuesta en color negro. En la figura 5.16 puede verse la gráfica generada al correr esta rutina con los con 81 espectros de emisión de fluorescencia analizados en la sección anterior (37 corresponden a orina de personas sanas y 44 a muestras de personas diagnosticadas con alguna patología oncológica), y un nivel de confianza ($1-\alpha$) del 99%.

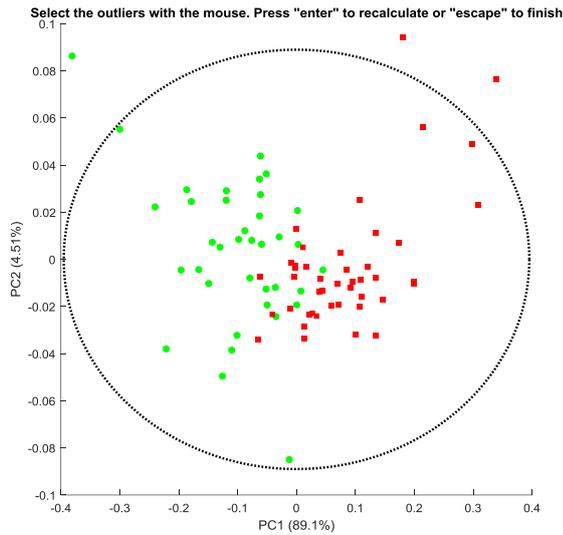


Figura 5.16 Puntos en el espacio PC1/PC2, correspondientes a las muestras de orina de personas sanas (en verde) y diagnosticadas con alguna patología oncológica (en rojo), y la elipse de confianza ($1-\alpha=99\%$) en color negro.

Luego, con el objetivo de eliminar las muestras muy poco representativas, se seleccionaron con el mouse los "outliers" o muestras que caen fuera de la elipse de la población total (Figura 5.17), que en este caso resultaron las muestras P001, P008 y N035.

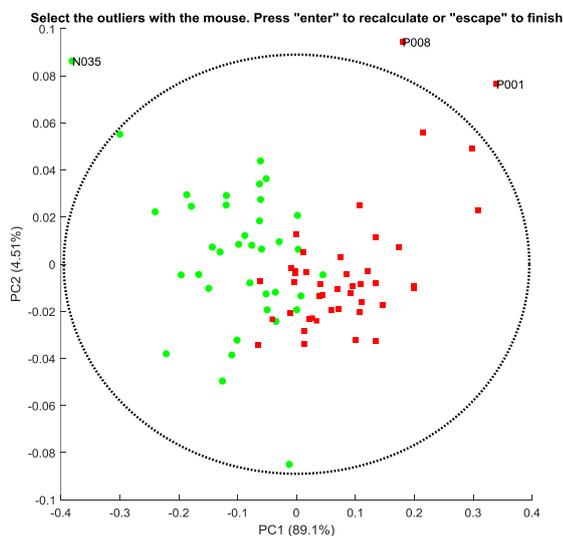


Figura 5.17 Selección de los outliers N035, P001 y P008, en el espacio PC1/PC2

Una vez seleccionados, al presionar "enter", la rutina se reinicia con la submatriz depurada y se recalculan tanto las proyecciones en el nuevo plano PC1/PC2 para el conjunto de datos reducido, como la nueva elipse sin considerar los outliers eliminados

(figura 5.18). Si se presiona "escape" el cálculo finaliza y la rutina genera una estructura de datos con los resultados de todos los cálculos.

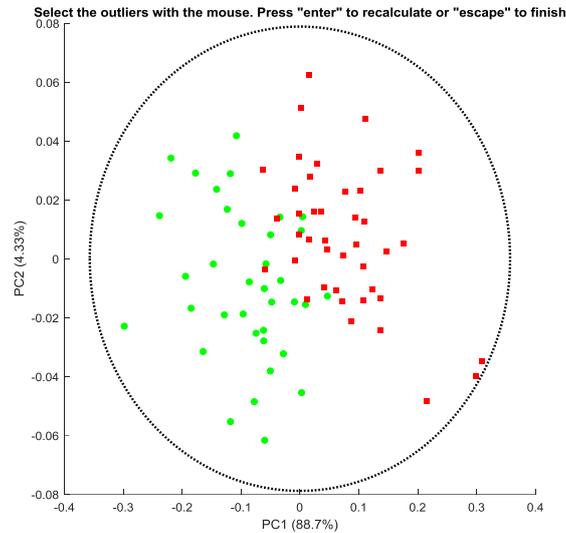


Figura 5.18 Puntos en el espacio PC1/PC2, y en color negro la elipse de confianza ($1-\alpha=99\%$), recalculada sin los outliers

Una vez eliminados los outliers de la población total, el software permite la generación de elipses independientes teniendo en cuenta por un lado sólo la población de enfermos y por el otro sólo la población de personas sanas. La gran ventaja que brinda tener una estimación independiente de las elipses correspondientes a cada población, es que queda desacoplada la correlación entre ambas poblaciones. En otras palabras, al considerar las poblaciones en forma independiente, los ejes de las elipses no están correlacionados y la variabilidad en cada dirección tampoco. De ahí se obtiene una distribución normal sobre cada eje, para cada población y por lo tanto el software calcula dos σ y dos μ para cada una de ellas.

La rutina grafica todos los puntos en el espacio PC1/PC2 con ambas elipses superpuestas. Los puntos y elipses que corresponden a muestras de personas clasificadas como "sanas" se grafican en color verde, y los correspondientes a personas "afectadas" en rojo. Esta es una manera muy clara de observar la distribución de puntos e identificar ambas poblaciones a simple vista (ver Figura 5.19).

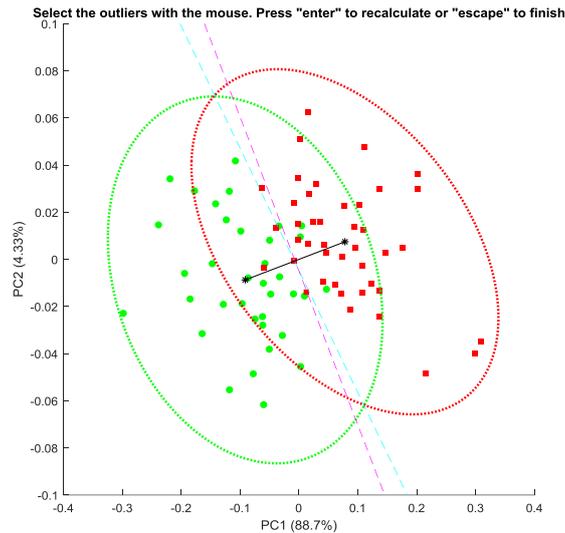


Figura 5.19 Puntos en el espacio PC1/PC2, con elipses de confianza ($1-\alpha=99\%$), calculadas a partir de las muestras de orina de personas sanas (en verde) y de las diagnosticadas con alguna patología oncológica (en rojo).

Para la elaboración de Tests Diagnósticos sobre el plano PC1/PC2, se pensaron y probaron varios criterios de separación de poblaciones. De la variedad analizada surgen como métodos de interés dos que se describen a continuación. Cada uno de ellos está basado en una recta que puede usarse como "umbral" para decidir si cada score es clasificado como positivo o negativo, según esté a un lado u otro de la misma. De esta manera se generaron dos test con diferentes criterios para el valor de corte:

1_“Ojiva”: El criterio de corte fue la línea que une los puntos de intersección de ambas elipses (o una paralela a ésta, como se verá a continuación).

2_“Centroide”: El criterio de corte estuvo dado por la mediatriz del segmento que une los centroides de las dos elipses entre sí (o alguna recta paralela a ésta).

En la figura 5.19 puede observarse que, además de los puntos correspondientes a cada muestra y las elipses de confianza, el software muestra el segmento que une los centroides de las elipses y las dos líneas punteadas que pueden usarse como "umbrales": la “ojiva” en color cyan, y el “centroide” en color magenta. El software está diseñado de manera que ambas líneas de corte puedan desplazarse a los lados, para poder evaluar ambos test con diferentes valores de corte. De esta manera, la cantidad de puntos que quedan a cada lado de la línea varía, cambiando la clasificación de éstos como correspondientes a personas “sanas” o “enfermas”, según el test. La figura 5.20 muestra ejemplos de los mismos datos, con las líneas de corte desplazadas hacia la derecha, y la figura 5.21 hacia la izquierda.

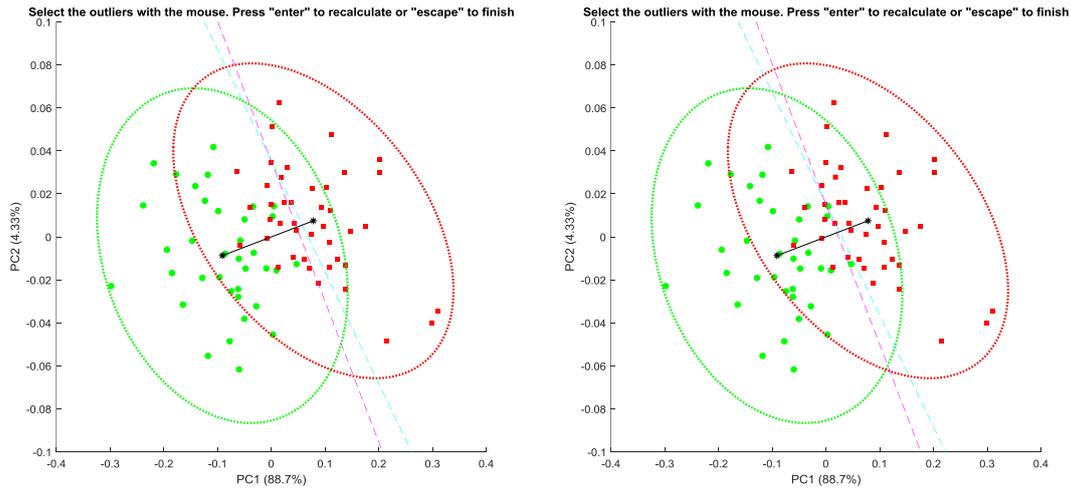


Figura 5.20: Ejemplos de líneas de corte desplazadas a la derecha: desplazamiento de 1 unidad (Izq.), y desplazamiento de 2 unidades (Der.).

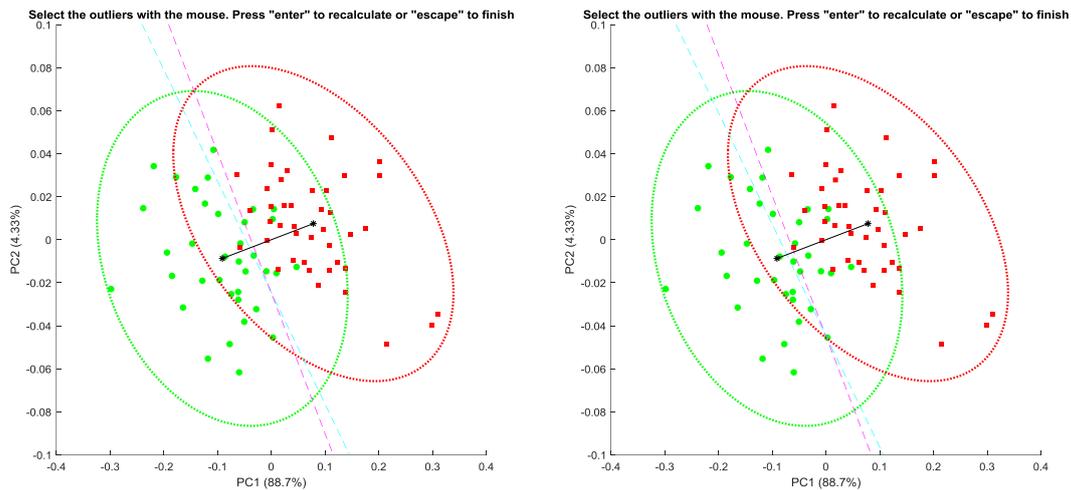


Figura 5.21: Ejemplos de líneas de corte desplazadas a la izquierda: desplazamiento de 1 unidad (Izq.), y desplazamiento de 2 unidades (Der.).

Con el objetivo de desarrollar Tests diagnósticos basados en los criterios mencionados, el software descrito fue adaptado para realizar las correspondientes clasificaciones comparando con las obtenidas según el estándar de oro (VP, VN, FP, FN). En este contexto, se y recalcularon automáticamente los indicadores de la validez del test (sensibilidad, especificidad y precisión) para una cierta cantidad de desplazamientos de las líneas de corte, y se generaron las curvas ROC correspondientes a cada test.

Con el fin de realizar una comparación de la capacidad discriminatoria de ambos tests se graficaron en una misma figura ambas curvas ROC superpuestas. Esto se realizó sucesivas veces, para diferente cantidad de intervalos, o corrimientos de las líneas umbral,

con el fin de evaluar si la capacidad discriminatoria de cada test variaba con ese parámetro. Entonces, cada curva ROC se generó a partir de un total de N corrimientos de la línea umbral, mitad a izquierda y mitad a derecha, con N variando entre 5 y 90. A modo de ejemplo, en la figura 5.22 se muestran las curvas obtenidas para N=10 y N=50. Allí se representaron en color azul las curvas ROC correspondientes al test llamado “Ojiva”, (líneas punteadas color cyan en las figuras 5.19 a 5.21), y en color verde las curvas ROC correspondientes al test denominado “Centroide” (líneas punteada color magenta en las mismas figuras).

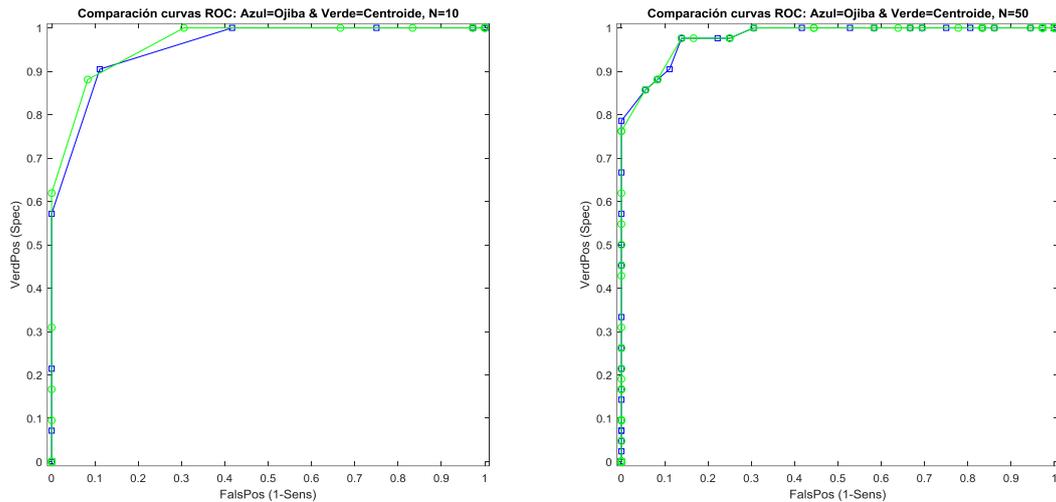


Figura 5.22: Comparación de las curvas ROC correspondientes los test “Ojiva” (en azul) y “Centroide” (en verde) realizadas con 10 (Izq.) y 50 desplazamiento de la línea umbral (Der.).

En la figura 5.22 (Izq.) parece haber una sutil diferencia entre las curvas ROC realizadas con pocos desplazamientos de las líneas umbral (valores pequeños de N). Sin embargo, ésta puede depreciarse para valores de N mayores a 30. Es decir, las curvas ROC realizadas con muchos puntos son prácticamente iguales, resultando obviamente en idénticos valores de AUC para ellas. Por lo tanto, a partir de las curvas ROC obtenidas, puede concluirse que ambos test poseen rendimientos similares en cuanto a la capacidad para discriminar entre muestras correspondientes a personas sanas y enfermas. Teniendo en cuenta que el valor de N puede afectar los valores de AUC de aquí en adelante, para todas las comparaciones entre diferentes tests que se muestren en el presente trabajo se compararán curvas ROC con la misma cantidad de puntos (N= 500).

A partir de los datos transformados por PCA, se estudió la posibilidad de encontrar algún otro test que pudiera mejorar los vistos anteriormente. En ese sentido se evaluaron, además de las dos líneas umbral, otros criterios de separación, y se valoró la capacidad de cada uno para discriminar entre ambas poblaciones. Tal como se explicó en la sección 4.3.1, PCA transforma los datos matriz original a partir de la construcción de unos pocos

ejes abstractos (PCs) que explican la variabilidad de los mismos. Para ello hace una descomposición de dicha matriz en dos matrices T y P de menor dimensión. Sabiendo que los componentes principales que arroja el PCA están ordenados en orden decreciente de porcentaje de variabilidad explicada, y teniendo en cuenta que este valor corresponde al 88,7% de la variación total para el PC1, y 4,33% para el PC2, se podría pensar que el score o proyección de cada muestra sobre el PC1 puede resultar un buen índice para el desarrollo de un test. Para indagar al respecto se realizó el análisis estadístico sobre la primera y segunda columna de la matriz T (scores PC1 y PC2, respectivamente), y se compararon los resultados. La figura 5.23 muestra los valores de los scores para PC1 y PC2, respectivamente. Del mismo modo que para las figuras de los índices de la sección 5.2, se graficaron superpuestos los scores obtenidos para personas sanas (en color verde) y los correspondientes personas afectadas (en rojo) para facilitar su visualización.

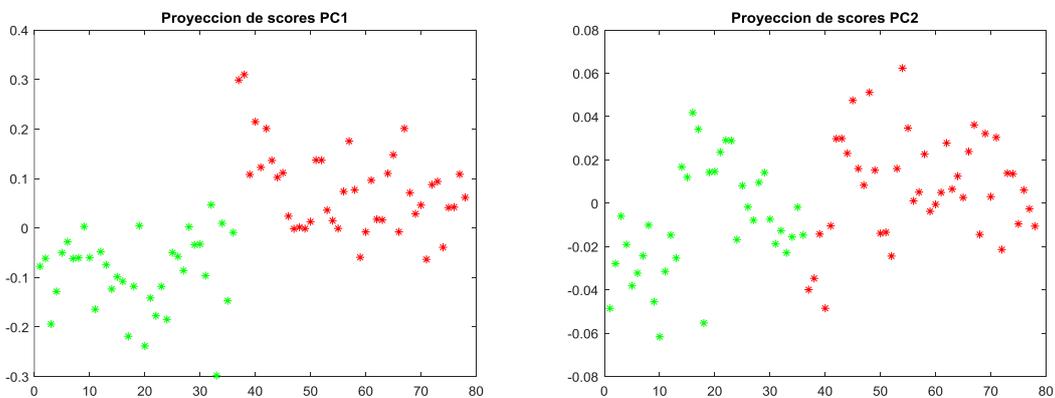


Figura 5.23 Proyecciones de los scores sobre PC1 (Izq.) y PC2 (Der.)

A partir del análisis de estas figuras puede observarse que en general los valores de los scores del PC1 para personas afectadas parecen ser mayores que los de las personas sanas, mientras que en los valores del PC2 no se observa ningún comportamiento relevante. Los histogramas correspondientes y las curvas ROC se presentan en las figuras 5.24 y 5.25.

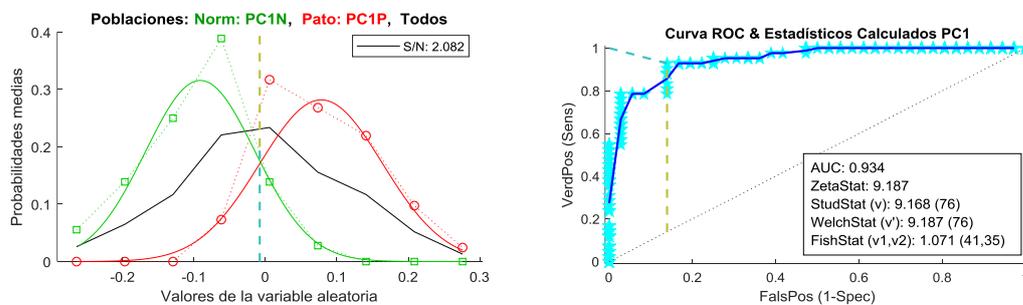


Figura 5.24 Histograma (Izq.) y curva ROC (Der.) correspondientes a los scores sobre el PC1, luego de la eliminación de los outliers.

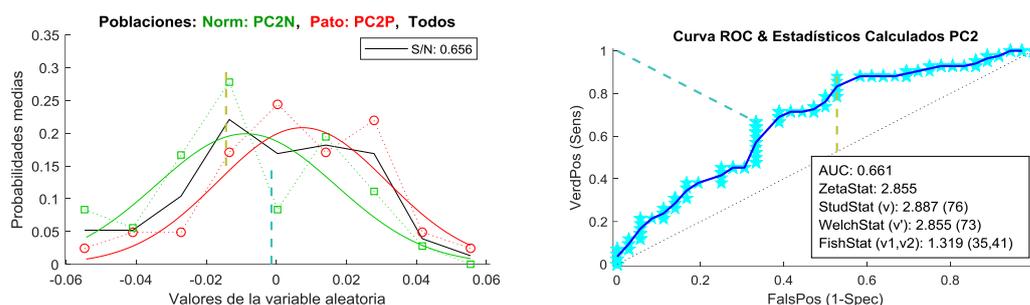


Figura 5.25 Histograma (Izq.) y curva ROC (Der.) correspondientes a los scores sobre el PC2, luego de la eliminación de los outliers.

Tal como se observa en las figuras, los valores obtenidos para la estimación de la relación señal ruido fueron 2,082 y 0,656 para los scores del primer y segundo componente principal, respectivamente. Dado que éste último valor es muy inferior a 1, se descartaron los scores del PC2 como índices de interés. Por otra parte, al analizar los scores del PC1, el estadístico F para las varianzas presentó un valor $F=1,071 < F_{\text{crítico}} (F^{0,01}_{41,35}=2,190)$, por lo que se asumió la homogeneidad de las varianzas, y se evaluó la diferencia entre las medias poblacionales con el test de Student. El t-valor obtenido fue 9,168, mucho mayor al $t_{\text{crítico}} (t^{0,01}_{76}=2,381)$, por lo que puede afirmarse que existe una diferencia significativa entre las medias de este índice para sanos y enfermos.

Finalmente, cabe destacar que el análisis de las curvas ROC mostró que el test generado a partir de los scores del PC1 superó la capacidad dicriminatoria de los test “ojiva” y “centroide” vistos anteriormente.

5.4.1 Estudio de la variabilidad intra-individuo

Tal como se explicó en el Capítulo 1, la composición de la orina de una persona puede variar dependiendo de diversos factores además de su estado de salud/enfermedad. Algunos de ellos son la cantidad de agua ingerida, el tiempo de retención, la dieta de la persona, etc. Estas variaciones en la composición pueden reflejarse en sus espectros de fluorescencia y por eso un aspecto muy importante a considerar a la hora de emplear valores de fluorescencia para la elaboración de un test diagnóstico es la variabilidad intra-paciente. Es decir, resulta necesario tener en cuenta las fluctuaciones que pueden presentar los espectros de las muestras de orinas de una misma persona (mismo estado de salud/enfermedad), tomadas en diferentes condiciones: diferentes días, diferentes horarios, diferentes tiempos de retención/ayuno, etc. Naturalmente, para el desarrollo de un test robusto, la variabilidad intra-paciente debe

compararse con la variabilidad inter-paciente . La presente sección detalla el empleo del PCA para el estudio de la variabilidad intra-paciente correspondiente a un individuo sano.

Es importante aclarar que el PCA fue el método seleccionado para este tipo de análisis ya que permitió hacer un “colapso” de las 138 dimensiones a sólo dos, reteniendo el 93,5% de la información total de los datos originales. Además, como consecuencia de esa reducción a dos dimensiones, el PCA permitió una evaluación gráfica sencilla de los resultados, tal como se muestra a continuación. En este contexto, se eligió esta poderosa y robusta herramienta matemática para realizar el estudio de la variabilidad intra-individuo, asumiendo que la conclusión que resulte de este análisis será válida para el resto del presente trabajo de Tesis.

Para analizar la variabilidad intra-individuo, a una misma persona perteneciente a la población de personas sanas le fueron solicitadas varias muestras de orina. Así, se recolectaron en diferentes días ocho muestras de la primer orina de la mañana en ayunas y con un tiempo de retención mayor a 5 hs. Además, se recolectaron otras cuatro muestras sin ayuno previo, en diferentes horarios: después de almorzar, antes de acostarse, durante la tarde con poco tiempo de retención, etc. En la tabla 5.4 se detallan las condiciones del momento de colección de cada muestra.

Fecha	Hora	Hs de Ayuno	Hs de retención
29/02/2016	7 AM	9	8
04/11/2016	7 AM	8	5,5
14/11/2017	7 AM	9	8
01/09/2017	7 AM	7	7
18/09/2017	7 AM	9	8
13/08/2019	7 AM	8	6
20/08/2019	7 AM	8	8
20/08/2019	4 PM	3	1
20/08/2019	6 PM	1	3
20/08/2019	3 AM	4	4
23/08/2019	7 AM	8	8
23/08/2019	11 PM	0,5	5

Tabla 5.4 Características de las muestras usadas para analizar la variabilidad intra-muestra

De esta manera se obtuvo un conjunto de 12 espectros de la misma persona, que se muestran en la figura 5.26.

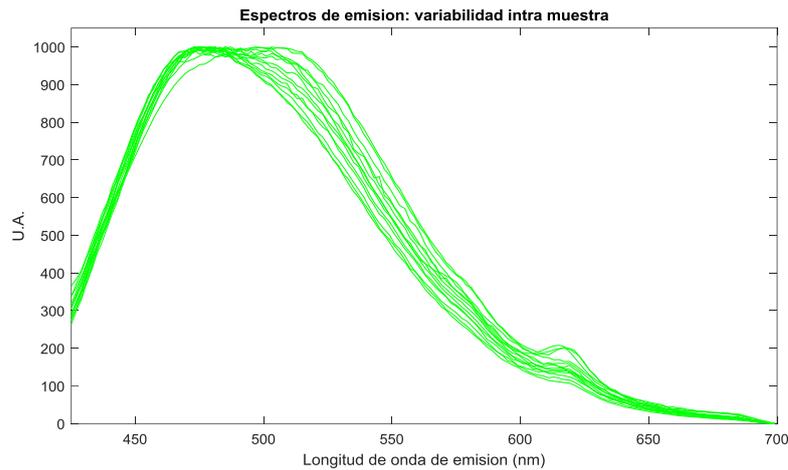


Figura 5.26 Espectros de emisión de diferentes muestras de orina de la misma persona.

Si bien todos los espectros son muy parecidos entre sí, a simple vista se observa que no son exactamente iguales. Los espectros presentan pequeñas diferencias, que se localizan principalmente entre 465 y 650 nm.

Con el fin de realizar una comparación más precisa que la simple inspección visual de los espectros, y poder cuantificar de alguna manera su variabilidad para compararla respecto de la variabilidad total entre muestras de diferentes individuos, se realizó una reducción de la dimensionalidad de la nueva matriz de doce espectros (\mathbf{X}') empleando los resultados de análisis PCA obtenido previamente con la matriz general \mathbf{X} . En otras palabras, no se realizó PCA empleando los espectros de la matriz \mathbf{X}' , sino que se proyectaron estos nuevos doce espectros sobre el plano PC1/PC2 previamente calculado con las 81 muestras originales \mathbf{X} . De esa manera se logró “visualizar” a los puntos correspondientes a la misma persona, en el mismo plano que el correspondiente a la población total de muestras analizada en las secciones anteriores, de manera de poder comparar sus dispersiones.

Para poder proyectar los nuevos puntos en el plano PC1/PC2 ya calculado, fue necesario encontrar las coordenadas (scores) de cada uno de los nuevos espectros contenidos en \mathbf{X}' en el plano original. El razonamiento utilizado para calcular las proyecciones se detalla a continuación:

La descomposición de la matriz de datos \mathbf{X} que hace el PCA se pueda representar matemáticamente como:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^T \tag{5.1}$$

donde \mathbf{X} es la matriz de datos original, \mathbf{T} los scores y \mathbf{P} los loadings. Multiplicando a derecha por \mathbf{P} a ambos lados de la igualdad, resulta:

$$\mathbf{X} \cdot \mathbf{P} = \mathbf{T} \cdot \mathbf{P}^T \cdot \mathbf{P} = \mathbf{T} \cdot \mathbf{I} \tag{5.2}$$

siendo I la matriz identidad. Cabe aclarar que la última igualdad es posible, debido a que la matriz real P es ortogonal, y por lo tanto cumple con la condición de que su inversa y su traspuesta son iguales. La expresión anterior puede escribirse en forma compacta como

$$T = X \cdot P \quad (5.3)$$

Por lo tanto, para calcular los scores T' correspondientes a los nuevos espectros X' , en principio podría usarse la ecuación 5.3:

$$T' = X' \cdot P, \quad (5.4)$$

donde P son los loadings que definen el plano PC1/PC2 calculados según la ecuación 5.1 a partir del conjunto de datos original. Sin embargo, dado que la rutina PCA antes de generar los loadings (P) y los scores (T), le resta a cada espectro de la matriz X el espectro promedio (covarianza alrededor de la media), para obtener las coordenadas finales de las 12 nuevas muestras sobre el plano generado anteriormente es necesario un paso adicional que se describe a continuación:

Se calculan los scores del espectro promedio de la población inicial X , utilizando la ecuación 5.3:

$$T_{Prom} = X_{Prom} \cdot P, \quad (5.5)$$

donde X_{Prom} es el espectro promedio del conjunto original de datos X . Las coordenadas finales para las 12 muestras nuevas (T'_{Final}) se obtienen entonces, a partir de la diferencia entre los scores obtenidos para las nuevas muestras (T') y los scores del espectro promedio (T_{Prom}):

$$T'_{Final} = T' - T_{Prom} \quad (5.6)$$

Una vez obtenidas estas nuevas coordenadas, cada espectro T'_{final} fue proyectado en el plano PC1/PC2 calculado a partir de la matriz X (81 muestras sanos y patológicos originales). En la figura 5.27 se observa la proyección, en el plano PC1/PC2 original, de los 37 espectros correspondientes a las muestras de personas sanas (en verde) y sobre ellos superpuestos los puntos correspondientes a los nuevos espectros X' de diferentes muestras de la misma persona. En color azul se muestran los puntos correspondientes a muestras de la primer orina de la mañana, y en magenta las que fueron colectadas en otros momentos del día, sin estar en ayunas y con tiempos de retención variables.

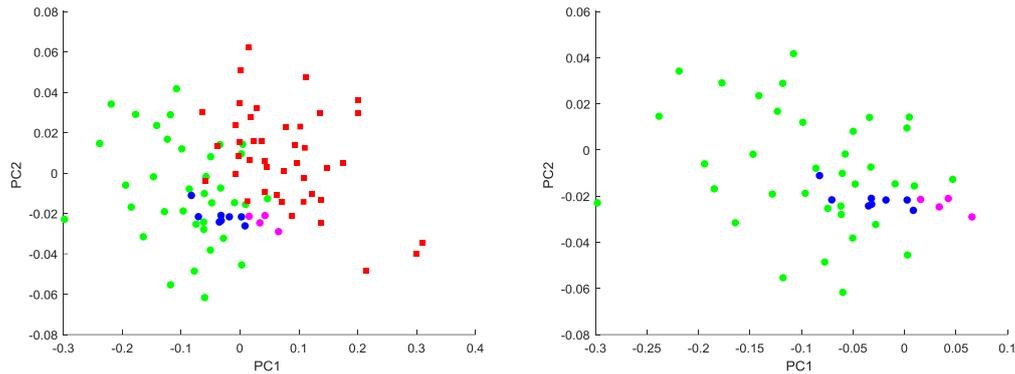


Figura 5.27 Proyección en el plano PC1/PC2 de los espectros correspondientes a las muestras de personas sanas (en verde), y a los espectros X' (en azul de la primer orina de la mañana, y en magenta sin ayuno y con tiempo de retención variable).

Tal como puede apreciarse, existe una variabilidad asociada a las condiciones de toma de la muestra, que se manifiesta en sus espectros de fluorescencia. Sin embargo, esta variabilidad “intra-individuo” es mucho menor a la variabilidad total “inter-individuo” tanto para la población total como para la de personas sanas en este caso.

Es importante destacar aquí, que las muestras del mismo paciente (X'), abarcan un período de 3 años, dando solidez a los resultados obtenidos. Es decir, a lo largo de los 3 años que transcurrieron entre la recolección de la primera y la última muestra, la variabilidad de los espectros de la orina de la misma persona fue sustancialmente menor a la variabilidad entre individuos de la población total.

También, en la figura 5.27 se observa que si bien la variabilidad intra-individuo es pequeña respecto de la total, los puntos correspondientes a las muestras que no fueron colectadas de la primer orina de la mañana (puntos magenta), son los que más contribuyen a incrementar este valor ya que se separan levemente de los azules, que corresponden a muestras tomadas en ayunas, e influyen aumentando la dispersión. Es decir, dos puntos correspondientes a la primer orina de la mañana, recolectada dos días distintos con 3 años de diferencia, se encuentran más cercanos entre sí en el plano PC1/PC2 que dos puntos correspondientes al mismo día pero en diferente estado de ayuno y retención. Este es un resultado sumamente importante a tener en cuenta en todo lo que resta del presente trabajo de Tesis, y será considerado en las conclusiones generales.

Resumiendo, a partir del análisis presentado en esta sección surgen dos conclusiones muy importantes:

1) A pesar de que la composición de la orina varía permanentemente, la variabilidad de la fluorescencia intra-paciente es mucho menor a la variabilidad total inter-paciente, por lo que éste método de test diagnóstico puede ser viable.

2) Para minimizar la variabilidad intra-paciente, y mejorar el rendimiento del test, es necesario analizar siempre la primer orina de la mañana, en ayunas y con suficiente tiempo de retención.

5.5 Análisis de los espectros normalizados y sus derivadas

Con el objetivo de encontrar otras características de los espectros que permitan discriminar desde el punto de vista estadístico las poblaciones de sanos y enfermos, se analizaron diferentes pre-procesamientos de los datos espectrales para ver cuál de ellos resultaba el más efectivo en la discriminación. En la presente sección se estudió la utilización del valor de intensidad de fluorescencia de emisión a cada longitud de onda como índice, para los espectros completos, y para sus derivadas primera y segunda. La primera derivada de un espectro (respecto de su longitud de onda), es una representación de la pendiente del espectro que da cuenta de la velocidad de cambio de la emisión de fluorescencia con la longitud de onda, mientras que la segunda derivada es una medida de la curvatura del espectro para una región de longitudes de onda considerada. De esta manera, la función derivada puede ser usada para resaltar pequeñas variaciones de los espectros.

Se trabajó con los mismos 81 espectros analizados anteriormente, pero en este caso adimensionados mediante el uso de normalización euclídea (ver sección 4.1.3). En la figura 5.28 (Sup.) se muestran los espectros normalizados: en color verde los correspondientes a personas sanas y en color rojo los que corresponden a orina de personas afectadas.

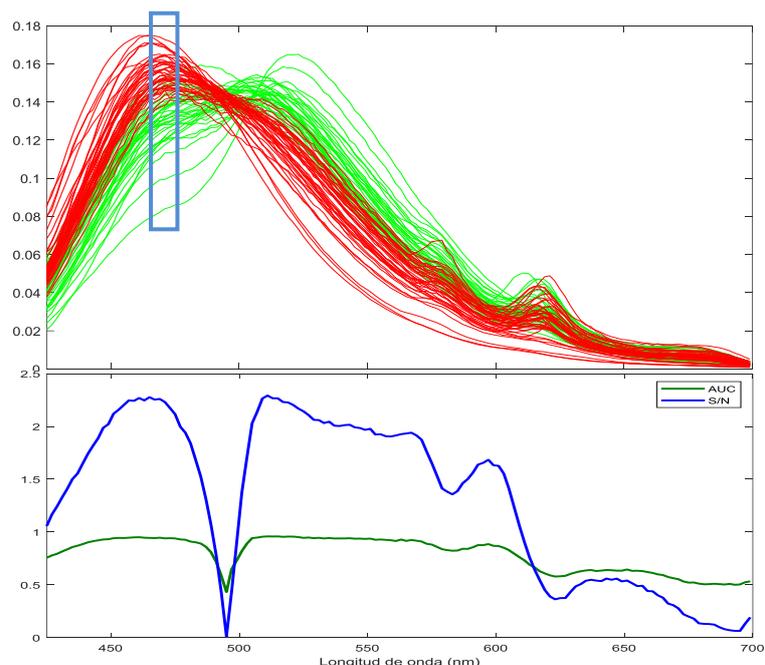


Figura 5.28 (Sup.) EEm de las muestras de orinas de personas sanas (en verde), y con patologías oncológicas (en rojo), según norma euclideana. (Inf.) Valores de AUC y S/N obtenidos.

Haciendo uso de las ventajas de la adimensionalización de los espectros dadas por la normalización, se estudiaron como índices de interés los valores de la emisión de fluorescencia a cada longitud de onda individual. De esta manera se analizaron 138 “índices” para los espectros sin derivar. Tal como puede observarse en la figura 5.28, la mayor parte de los EEm, presentan valores superpuestos para las muestras correspondientes a sanos y enfermos (curvas verdes y rojas solapadas). Esto indicaría que la mayoría de las longitudes de onda de emisión no serían “buenos” índices para el desarrollo de tests diagnósticos. Sin embargo, existen algunas regiones, como la comprendida entre 455 y 465 nm por ejemplo (recuadro azul), que presentan poco solapamiento, por lo que podrían servir para discriminar las poblaciones. Para confirmar esa suposición, se realizó el análisis estadístico a todos los “índices”, es decir se analizó individualmente cada longitud de onda de emisión de todo el espectro. En la figura 5.28 (Inf.) se muestran los valores obtenidos para la relación señal/ruido (en azul), y el correspondiente valor del AUC para la curva ROC (en verde), de cada índice. Las figuras 5.28 Inf. y Sup. se realizaron con la misma escala en el eje de las abscisas y se dispusieron una sobre la otra, con el fin de facilitar su comparación e interpretación.

Tal como puede observarse, la relación S/N (curva azul) presenta dos máximos bien definidos en 465 nm (S/N=2,277) y 511 nm (S/N=2,293), y un máximo relativo en 597 nm (S/N=1,685), que a su vez se corresponden con los máximos valores del AUC (curva verde). Por otra parte los índices correspondientes a longitudes de onda de emisión entre 493 y 499nm, y de 613nm en adelante no aportan información en cuanto a la presencia de las dos poblaciones.

En consecuencia, de las 138 longitudes de onda que posee cada espectro de emisión, se eligieron como índices de interés las intensidades registradas a 465, 511 y 597 nm. A estos índices se les realizó el análisis estadístico, con la rutina ya descrita, y se obtuvieron los valores de los estadísticos que se listan en la tabla 5.5.

Derivada	Índice	S/N	F-Fisher	t-Student	t-Welch	(f)	AUC
Orden Cero	$I_{EuclI465}$	2,277	2,122(36,43)	10,079 (79)	9,772	(66)	0,945
Orden Cero	$I_{EuclI511}$	2,293	1,878(43,36)	-10,149 (79)	-10,428	(75)	0,959
Orden Cero	$I_{EuclI597}$	1,685	1,155(43,36)	-7,457 (79)	-7,513	(79)	0,886

Tabla 5.5 Valores de los estadísticos estimados para los índices I_{EuclI}

El estadístico F para las varianzas del índice 465 nm presentó un valor $F_{36,43}=2,122$, que es mayor al $F_{crítico}(F^{0,01}_{36,43}=2,118)$, por lo que en este caso no pudo asumirse la homogeneidad de las varianzas y se recurrió al test de Welch para la comparación de las medias. El valor estimado del estadístico $t=9,772$ fue mucho mayor al t de Student de 66 grados de libertad ($t^{0,01}_{76}=2,395$), por lo que se concluyó que, según este índice, existen diferencias significativas entre las medias para sanos y enfermos. Los valores estimados de los estadísticos F para las varianzas de los índices 511 y 597 nm (tabla 5.5) fueron menores al $F_{crítico}(F^{0,01}_{43,36}=2,16)$, por lo que en ambos casos se asumió la homogeneidad de las varianzas y se evaluó la diferencia entre las medias poblacionales con el test de Student. Dado que los valores t obtenidos fueron mucho mayores al valor $t_{crítico}(t^{0,01}_{79}=2,379)$, aquí también se asumió que existe una diferencia significativa entre las medias de estos dos índices para sanos y enfermos.

Teniendo en cuenta que los tres índices tienen la capacidad de discriminar ambas poblaciones, se compararon los valores de sus AUC, a partir de las curvas ROC correspondientes (figuras 5.29 a 5.31).

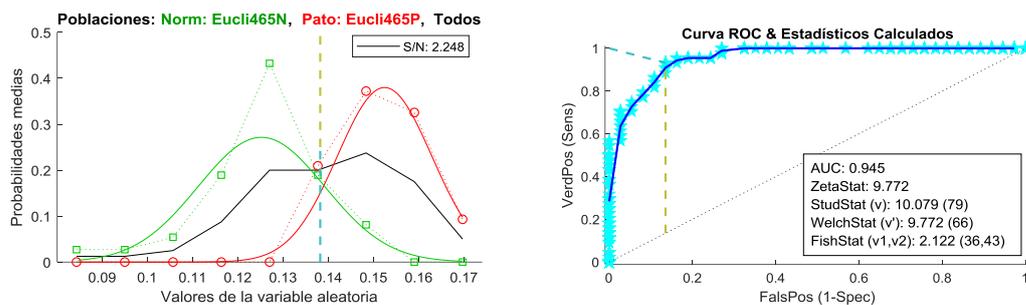


Figura 5.29 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice Eucli465

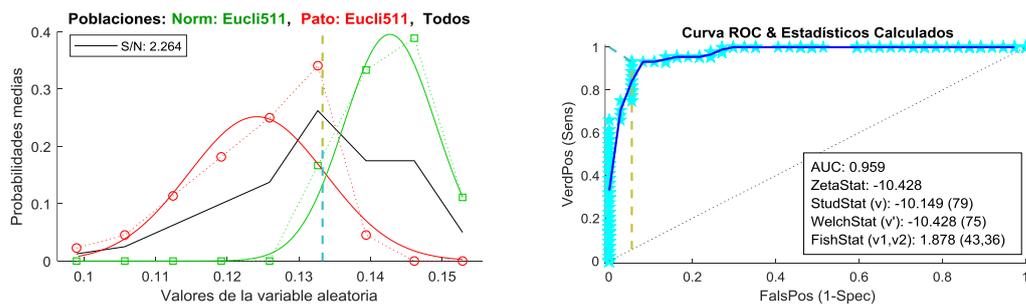


Figura 5.30 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice Eucli511

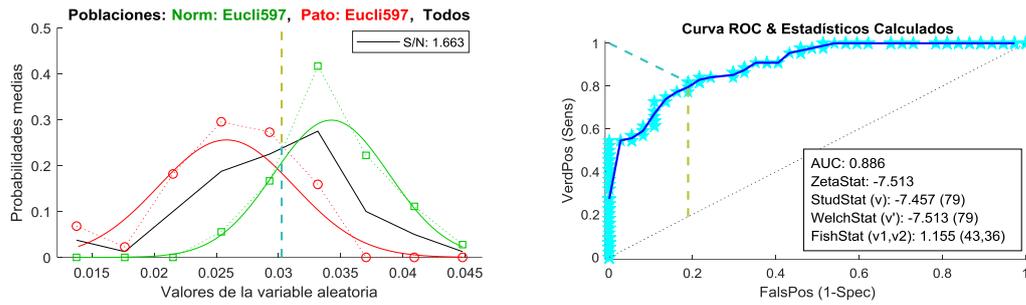


Figura 5.31 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice Eucli597

Los valores de AUC calculados fueron 0,945, 0,959 y 0,886 para las longitudes de onda 465, 511 y 597 nm, respectivamente. A partir del análisis expuesto se concluye que los índices de los espectros, escalados a norma euclídea, que mejor discriminan ambas poblaciones son las longitudes de onda de emisión a 465 y 511nm.

Por último, se calculó el índice $I_{Eucli465-511}$, como la diferencia de las intensidades (normalizadas) de emisión a 465 y a 511 nm, para compararlo con el ya calculado $I_{Area471-507}$ propuesto en la sección anterior como representativo de la diferencia de intensidad de las longitudes de onda correspondientes a los máximos de cada población. Los resultados del análisis estadístico realizado, los histogramas y la curva ROC correspondientes se muestran en la figura 5.32.

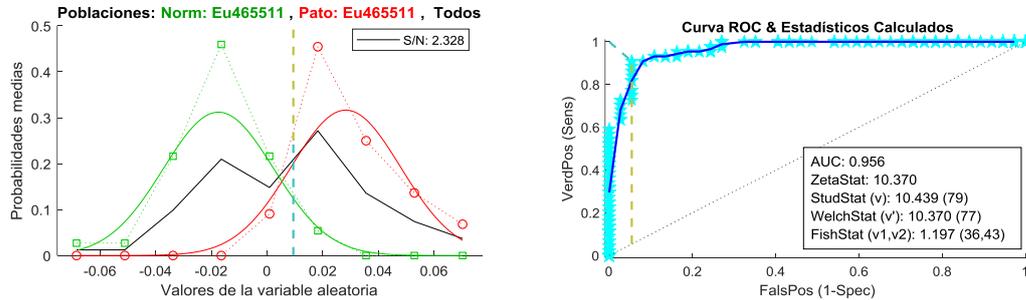


Figura 5.31 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice $I_{Eucli465-511}$.

El valor del estadístico F para las varianzas obtenido (1,197) fue menor al $F_{crítico}$ ($F^{0,01}_{36,43}=2,16$) por lo que se asumió la homogeneidad de las varianzas y se evaluó la diferencia entre las medias poblacionales con el test de Student. El valor t obtenido $t=10,439$ fue mucho mayor al valor $t_{crítico}$ ($t^{0,01}_{79} = 2,379$), por lo que se puede afirmar que existe una diferencia significativa entre las medias de este índice para pacientes sanos y enfermos ($p<0,0005$). Por lo tanto este índice también sirve para distinguir ambas poblaciones. La relación señal/ruido calculada para $I_{Eucli465-511}$ es comparable y levemente superior a la de $I_{Area471-507}$ (2,328 vs 2,300), mientras que el AUC aumentó en el mismo sentido (0,956 vs 0,952). Finalmente, usando el AUC como estimación del rendimiento de

los test se concluye que este nuevo índice presenta leves mejoras respecto del índice

|Area471-507·

Derivada de orden 1:

La técnica de derivar los espectros, conocida como “espectroscopia de derivadas”, es habitualmente empleada para mejorar la selectividad y el poder de resolución en problemas de interés bioquímico, clínico y farmacéutico, entre otros [78] La espectroscopia de derivadas permite resolver las bandas traslapadas (hombros) y puede facilitar la interpretación de ciertas características del espectro, reduciendo la interferencia del ruido de fondo (background).

Si bien esta metodología puede presentar muchas ventajas, es importante tener en cuenta las dificultades asociadas con la evaluación de derivadas por métodos numéricos a partir de intensidades de emisión experimentales (que obviamente constituyen un conjunto discreto). Cuando se deriva numéricamente un espectro registrado experimentalmente, la relación S/N puede impactar sustancialmente en la calidad de la estimación de la función derivada. El ruido tiene características muy agudas que afectan al espectro, y una estrategia para atenuar su efecto sobre las derivadas consiste en emplear una región de longitudes de onda relativamente ancha comparada con la estructura del ruido. Sin embargo, si bien la estimación de la derivada resulta más suave a medida que se incrementa el rango de longitudes de onda, la aproximación de orden cero a la función se aparta de los datos experimentales. Por lo tanto, existe un estrecho compromiso entre la relación S/N y el ancho del intervalo seleccionado para el análisis numérico. Tomando en consideración esta vinculación para calcular las derivadas de la presente sección se utilizó una rutina de Matlab, desarrollada previamente por el Dr. García Einschlag, que permite seleccionar manualmente el intervalo de longitudes de onda a analizar de manera de poder intervenir en este compromiso. Dicha rutina filtra los datos y calcula las derivadas primera y segunda en forma automática mediante el método de Savitzky-Golay.[79, 80]

Haciendo uso de la rutina mencionada se ensayaron diferentes tamaños de intervalos y ordenes de filtración. A partir de los resultados obtenidos se resolvió calcular las derivadas tomando siempre intervalo de 5 puntos a cada lado de la longitud de onda considerada y un orden 2 para los polinomios usados en el análisis de regresión. De esta manera, se obtuvieron valores para la derivada de los espectros en el rango de 437 a 689 nm. En la figura 5.32 se muestran las curvas obtenidas, siguiendo la convención de colores utilizada a lo largo de todo el trabajo: verde corresponde a las derivadas de los espectros de orina de personas sanas, y rojo a los de las patológicas.

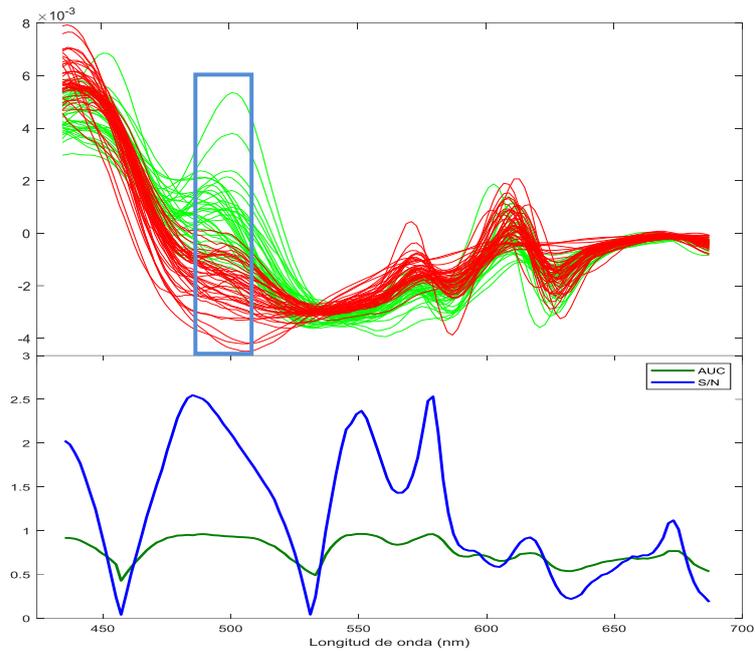


Figura 5.32 (Sup.) Primera derivada de los EEm de la figura 5.28.(Inf.) Valores de AUC y S/N obtenidos.

Los perfiles obtenidos sugieren que la zona comprendida entre 480 y 500 nm (recuadro azul) parece discriminar bastante bien ambas poblaciones y también muestran otras zonas que podrían resultar de interés. Para analizar todo el rango completo, se procedió de la misma manera que con los espectros sin derivar: se realizó el análisis estadístico a cada una de las longitudes de onda (“índices”), barriendo todo el espectro derivado, y luego se compararon los resultados. En el panel inferior de la figura 5.32 se muestran los valores obtenidos de la relación señal/ruido (en azul), y el correspondiente valor del AUC (en verde) calculado a partir de la curva ROC de cada índice. Allí puede observarse con claridad que la relación S/N presenta tres máximos bien pronunciados, y de valores superiores a los obtenidos a partir de los espectros sin derivar. Dichos máximos se encuentran en las siguientes longitudes de onda: 485-489 nm, 551 nm y 579 nm y se corresponden también con los mayores valores de AUC. El mayor valor de la relación S/N obtenido fue 2,515 y corresponde a la longitud de onda 579 nm, mientras que el AUC máximo fue de 0,963 y corresponde a la curva ROC de índice 551 nm. En las figuras 5.32 a 5.35, se muestran los histogramas y curvas ROC correspondientes a los índices I_{D1-485} , I_{D1-489} , I_{D1-551} e I_{D1-579} . El nombre de dichos índices fue asignado de manera que la letra “D” se refiere a “Derivada”, el “1” que está a continuación indica que es derivada de primer orden y los últimos tres dígitos se refieren a la longitud de onda considerada.

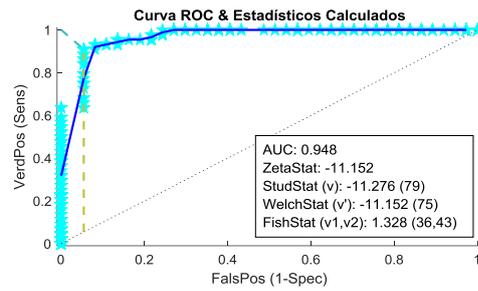
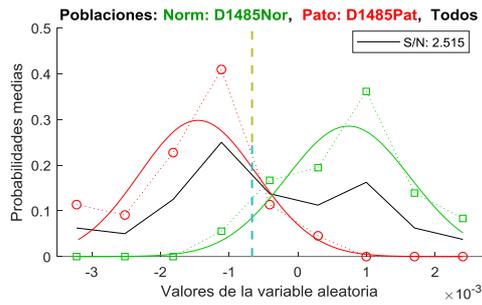


Figura 5.32 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice I_{D1-485} .

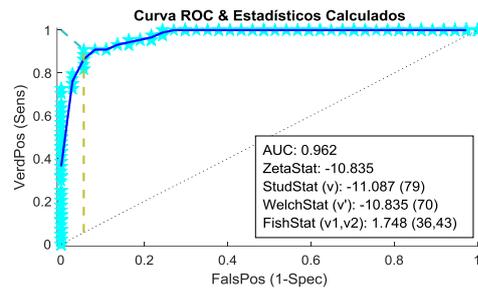
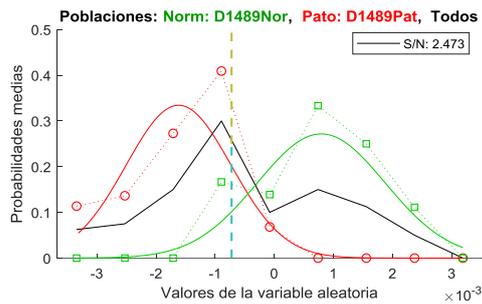


Figura 5.33 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice I_{D1-489} .

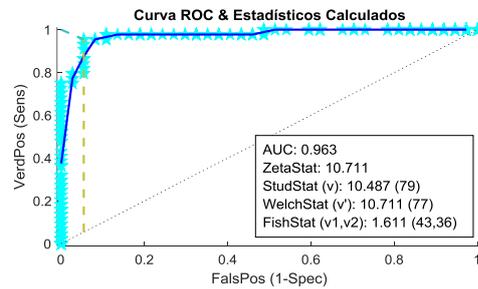
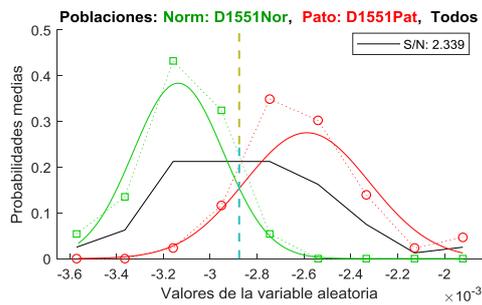


Figura 5.34 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice I_{D1-551} .

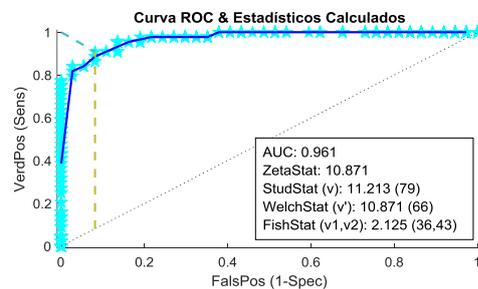
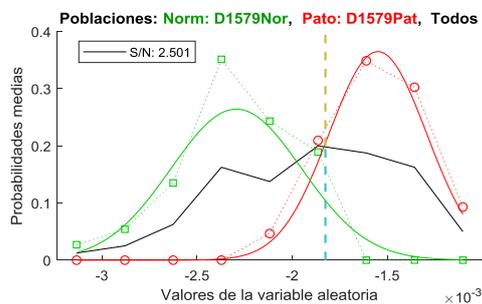


Figura 5.35 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice I_{D1-579} .

Los valores obtenidos del estadístico F para las varianzas de los índices I_{D1-485} , I_{D1-489} e I_{D1-551} resultaron inferiores (tabla 5.6, a continuación) a sus correspondientes $F_{\text{críticos}}$

($F^{0,01}_{36,43}=2,118$ y $F^{0,01}_{43,36}=2,16$), por lo que en los tres casos se asumió la homogeneidad de las varianzas y se evaluó la diferencia entre las medias poblacionales con el test de Student. Dado que en los tres casos los valores del estadístico t obtenidos (Tabla 5.6) fueron mucho mayores al t-crítico ($t^{0,01}_{79} = 2,379$), puede afirmarse que existen diferencias significativas entre las medias de estos índices para sanos y enfermos.

Derivada	Índice	S/N	F-Fisher	t-Student	t-Welch	(f)	AUC
1er orden	I _{D1-485}	2,515	1,328 (36,43)	-11,276 (79)	-11,152 (75)	(75)	0,948
1er orden	I _{D1-489}	2,473	1,748(36,43)	-11,087 (79)	-10,835 (70)	(70)	0,962
1er orden	I _{D1-551}	2,339	1,611(43,36)	10,487 (79)	10,711 (77)	(77)	0,963
1er orden	I _{D1-579}	2,501	2,125(36,43)	11,213 (79)	10,871 (66)	(66)	0,961

Tabla 5.6 Valores de los estadísticos estimados para los índices I_{D1}.

Para el índice I_{D1-579} el valor estimado del estadístico $F_{36,43}=2,125$, fue mayor que el F crítico correspondiente ($F^{0,01}_{36,43}=2,118$), por lo que se recurrió al test de Welch para la comparación de las medias. Dado que el valor del estadístico t=11,213 obtenido fue mucho mayor al t de Student de 66 grados de libertad, se concluye que también con este índice existen diferencias significativas entre las medias para sanos y enfermos.

Derivada de orden 2:

Del mismo modo que se analizaron los índices que surgieron a partir de las derivadas de primer orden, se calcularon los correspondientes a las derivadas de segundo orden. Empleando la misma rutina de Matlab descrita previamente, se calculó la segunda derivada de los espectros tomando también un intervalo de 5 puntos a cada lado. La figura 5.36 muestra las curvas obtenidas, y en ella puede observarse, además, el “incremento en el ruido” al aumentar el orden de la derivada considerada.

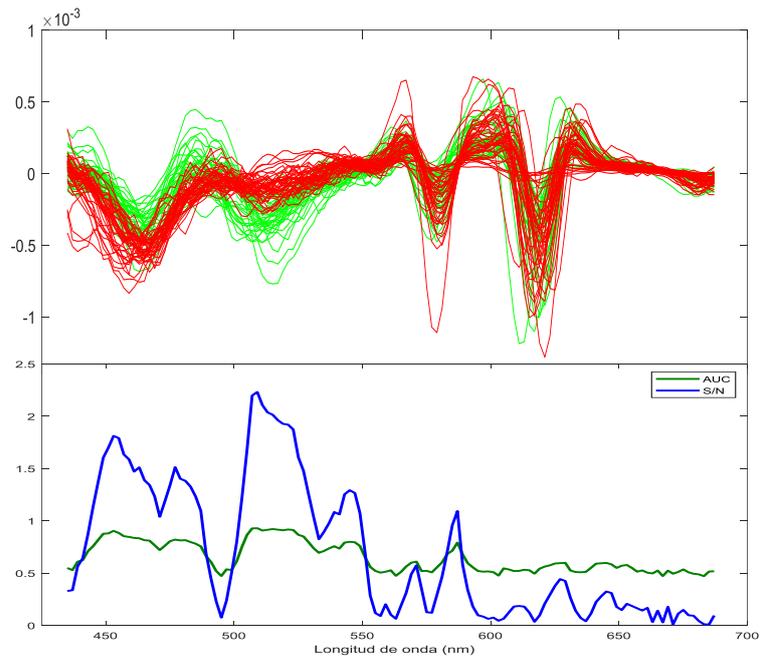


Figura 5.36 (arriba) Segunda derivada de los EEm de la figura 5.28.(abajo) Valores de AUC y S/N obtenidos.

En la figura 5.36 se muestran los valores de la relación S/N (azul) y del AUC (verde), para todas las longitudes de onda. De todo el rango de índices calculados, el único que se destaca y merece ser analizado es el que corresponde a la longitud de onda de emisión 509 nm, denominado I_{D2-509} . Dicho índice fue seleccionado por haber presentado los valores máximos de S/N y AUC ($S/R_{m\acute{a}x}=2,105$ y $AUC_{m\acute{a}x}=0,929$). Los valores de los estadísticos calculados para este índice mediante el software descrito se muestran en la tabla 5.7.

Derivada	Índice	S/N	F-Fisher	t-Student	t-Welch	(f)	AUC
Orden dos	I_{D2-509}	2,105	3,104(36,43)	9,435	(79)	9,017	(58) 0,929

Tabla 5.7 Valores de los estadísticos estimados para los índices I_{D2}

El valor del estadístico F para las varianzas de dicho índice $F=3,104$ (36,43) resultó mayor al $F_{crítico}$, por lo que en este caso no pudo asumirse la hipótesis de homogeneidad de las varianzas y se recurrió al test de Welch para la comparación de las medias. El valor del estadístico $t=9,017$ calculado fue mucho mayor al t de Student de 58 grados de libertad ($p<0,01$), por lo que se concluyó que también con este índice existen diferencias significativas entre las medias para sanos y enfermos, por lo que podría ser de utilidad como test diagnóstico. En la figura 5.37 se muestra el análisis poblacional y la correspondiente curva ROC.

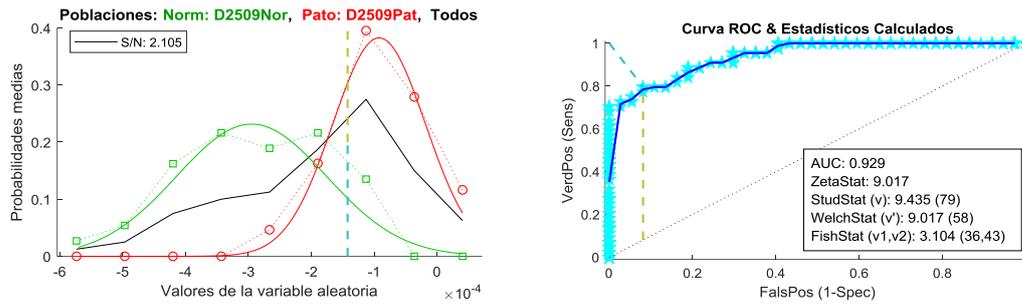


Figura 5.37 Histograma (Izq.) y curva ROC (Der.) correspondientes al índice I_{D2-509} .

5.6 Análisis por Resolución Multivariada de Curvas (MCR)

Otra de las técnicas multivariadas que se aplicó en el presente trabajo de Tesis para reducir la dimensionalidad de las MEE fue la resolución multivariada de curvas por cuadrados mínimos alternados (MCR-ALS). Tal como se explica en el Capítulo 4, la matriz de espectros medidos se descompuso en otras dos matrices C y S correspondientes a los scores (o pesos relativos), y a los perfiles espectrales, respectivamente. Cabe destacar que para lograr esto se asumió la validez del principio de superposición lineal, es decir que la fluorescencia total a cada longitud de onda es la suma de las contribuciones asociadas a la fluorescencia de cada componente del sistema. Por lo tanto la matriz C calculada contiene los coeficientes de la combinación lineal que, multiplicados por los perfiles espectrales de cada componente (S), permiten reproducir los espectros.

Se corrió el algoritmo MCR-ALS sobre los 81 espectros de emisión de fluorescencia medidos (37 correspondientes a personas sanas y 44 a personas afectadas por alguna patología oncológica) usando entre 2 y 7 componentes. El número máximo de iteraciones se fijó en 50, y se estimaron las contribuciones espectrales iniciales aplicando el procedimiento SIMPLISMA. En ambos modos (muestra y longitud de onda), la única restricción aplicada fue la de “no negatividad” para todas las deconvoluciones. Luego del proceso iterativo, con fines comparativos, los perfiles espectrales se normalizaron a la intensidad del máximo.

Para determinar el número más adecuado de componentes se evaluó la falta de ajuste de los datos (LOF) para diferente número de componentes. Los resultados obtenidos se muestran en la tabla 5.8.

Componentes	LOF
2	4.26
3	3.26
4	2.05
5	1.22
6	0.96
7	0.77

Tabla 5.8 Falta de ajuste (LOF) de los datos del análisis por MCR-ALS para 2 a 7 componentes

Tal como puede observarse, la falta de ajuste disminuye notablemente al aumentar la cantidad de componentes empleados, y esta disminución se vuelve menos significativa al incrementar el número de componentes. Es decir, el ajuste de los datos mejora más entre las componentes 2 a 3 ($LOF_3 - LOF_2 = 1$), que entre los componentes 6 y 7 ($LOF_7 - LOF_6 = 0,19$). Por el contrario, al aumentar el número de componentes por encima del valor necesario para reproducir correctamente la matriz experimental, los perfiles espectrales se deforman perdiendo de esta manera significado físico. Las figuras 5.38 a 5.40 muestran, a modo de ejemplo, los perfiles espectrales para los ajustes obtenidos por MCR con 5, 4 y 3 componentes. El análisis de los resultados revela que existe un compromiso entre la calidad del ajuste y la forma de los espectros obtenidos para diferente cantidad de componentes.

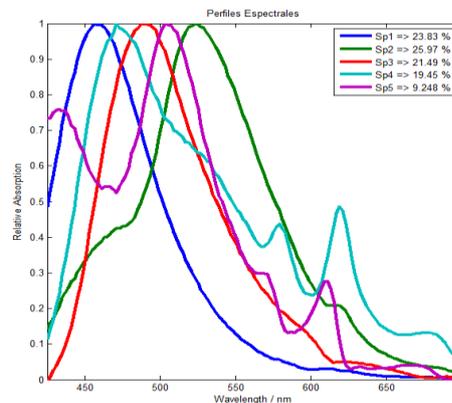


Figura 5.38 Perfiles espectrales del ajuste por MCR-ALS con 5 componentes

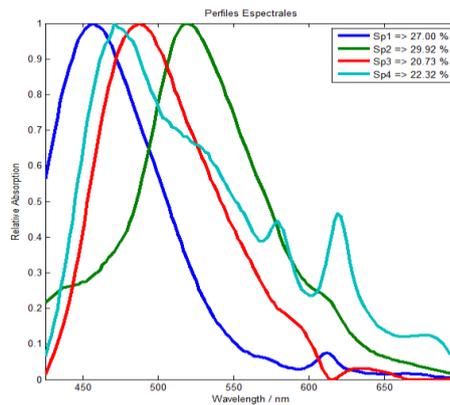


Figura 5.39 Perfiles espectrales del ajuste por MCR-ALS con 4 componentes

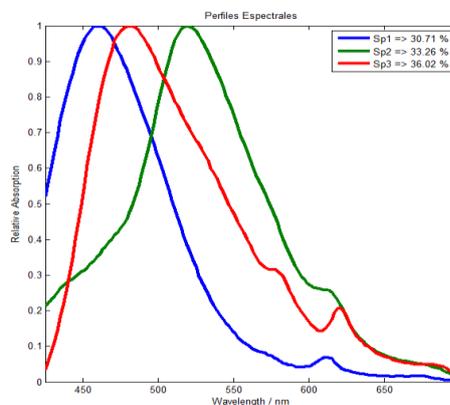


Figura 5.40 Perfiles espectrales del ajuste por MCR-ALS con 3 componentes

De todo lo expuesto, se consideró que 4 era el número de componentes óptimo para realizar el ajuste en este caso. Por lo tanto, de aquí en adelante, se presentará el análisis de la distribución de los coeficientes de la combinación lineal (scores) que permiten reproducir los espectros a partir del ajuste por MCR-ALS con 4 contribuciones. El valor de dichos scores se representa en los ejes de las ordenadas de la figura 5.41, en función del número de muestra (en las abscisas).

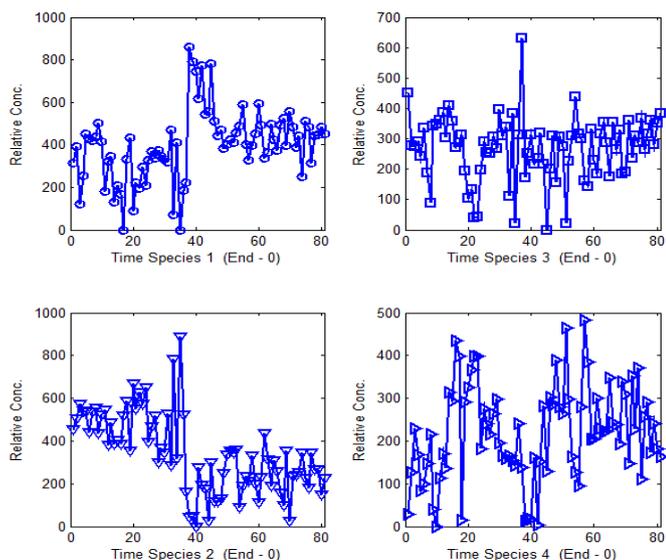


Figura 5.41 Scores del ajuste por MCR-ALS con 4 componentes. El cuadro superior izquierdo corresponde a los scores del componente 1, el inferior izquierdo al componente 2, el superior derecho a los del componente 3, y el inferior derecho al componente 4.

En la figura se observa que las componentes 1 y 2 parecen reflejar la presencia de las dos poblaciones. Es decir, la componente 1 presenta scores de menor valor para las muestras correspondientes a personas sanas (1 a 37) respecto de los scores de las muestras de personas afectadas (38 a 81), mientras que para la componente 2 se observa lo opuesto. En cambio, en los gráficos de los scores de las componentes 3 y 4 no se evidencia ningún comportamiento que pueda ser empleado para la discriminación entre poblaciones.

Para cuantificar estas observaciones, se evaluó el rendimiento de cada componente como índice de discriminación de poblaciones mediante el análisis estadístico descrito en secciones anteriores. Los resultados obtenidos sobre cada uno de los 4 grupos de scores se muestran en las figuras 5.42 a 5.45.

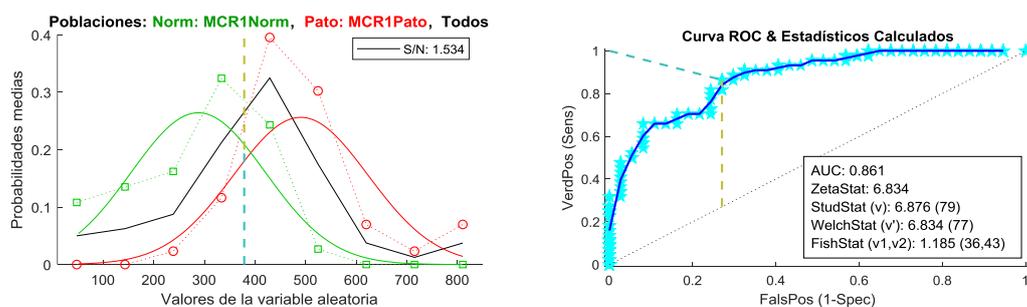


Figura 5.42 Histograma (Izq.) y curva ROC (Der.) correspondientes al Componente1 MCR

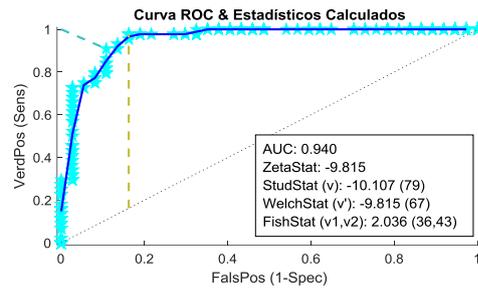
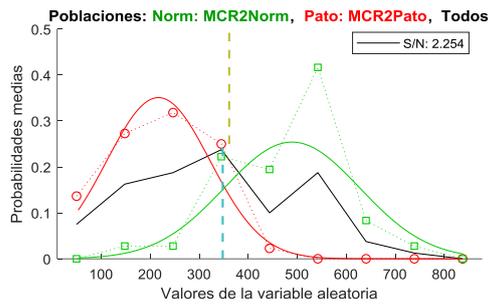


Figura 5.43 Histograma (Izq.) y curva ROC (Der.) correspondientes al Componente 2 MCR

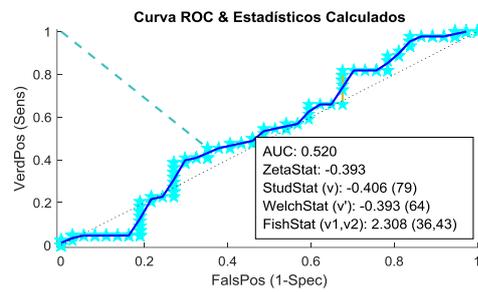
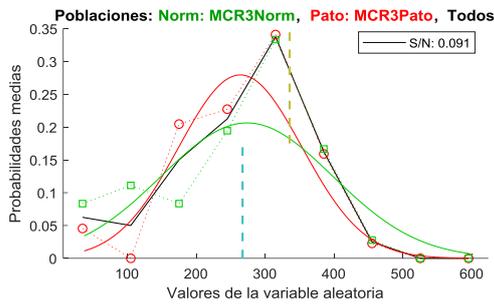


Figura 5.44 Histograma (Izq.) y curva ROC (Der.) correspondientes al Componente 3 MCR

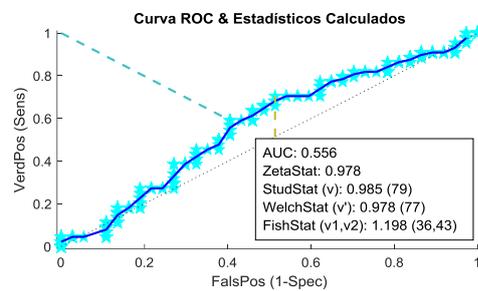
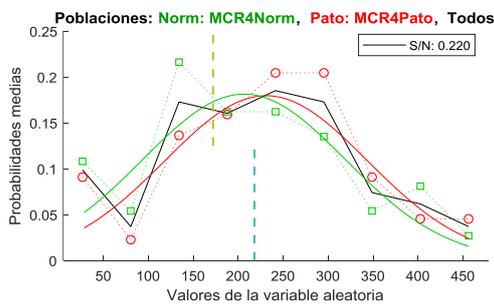


Figura 5.45 Histograma (Izq.) y curva ROC (Der.) correspondientes al Componente 4 MCR

Los valores calculados del estadístico F para los índices I_{MCR1} y I_{MCR2} (Figuras 5.42 y 5.43) fueron menores al valor crítico correspondiente ($F^{0,01}_{36,43}=2,118$), por lo que se asumió la homogeneidad de las varianzas, y se compararon los valores de las medias con el test de Student. En ambos casos los valores del estadístico t obtenidos fueron superiores a su correspondiente valor crítico de tabla ($t^{0,01}_{79} = 2,379$), por lo que se rechazó la hipótesis nula y se concluyó que existen diferencias significativas entre las medias poblacionales para ambos índices. Además de la valoración estadística, y en concordancia con lo que se observa en las figuras 5.42 a 5.45 (I_{MCR1} : AUC=0,861 y I_{MCR2} : AUC=0,940), se concluye que el índice I_{MCR2} es el que mayor poder de discriminación posee.

A partir del análisis de las figuras 5.44 y 5.45 se verifica lo observado en los gráficos de los scores (Figura 5.41): ninguno de estos dos índices (scores de las componentes 3 y 4) es útil a los fines de detectar la presencia de dos poblaciones. Brevemente, a partir de los valores estimados del estadístico F, se compararon las medias con el test de Welch para I_{MCR3} y con el test de Student para I_{MCR4} . En ambos casos los valores obtenidos del estadístico t ($t_{64}=0,393$ y $t_{79}=0,985$ para I_{MCR3} e I_{MCR4} , respectivamente) fueron menores a los valores críticos de tabla correspondientes. En consecuencia, en ambos casos se aceptó la hipótesis nula, de que las medias poblacionales son iguales.

Por otro lado, más allá de las evaluaciones de los componentes individuales, en la figura 5.41, se observó que los scores correspondientes a las muestras de personas sanas (1 a 37) presentaron, en general, valores inferiores a los de las personas afectadas (38 a 81) para el componente 1, y mostraron el comportamiento contrario para el componente 2. Esto mismo se manifiesta en los histogramas de dichas componentes (figuras 5.42 y 5.43), donde puede observarse que las posiciones de las curvas verde (sanos) y roja (patológicos) se encuentran invertidas. Entonces se evaluó como índice de interés el valor de la diferencia entre ambos scores, es decir, se generó un nuevo índice como la resta del valor del componente 1 menos el componente 2 para cada muestra. Los histogramas correspondientes, junto con la curva ROC y los valores de los estadísticos resultantes se muestran en la figura 5.46.

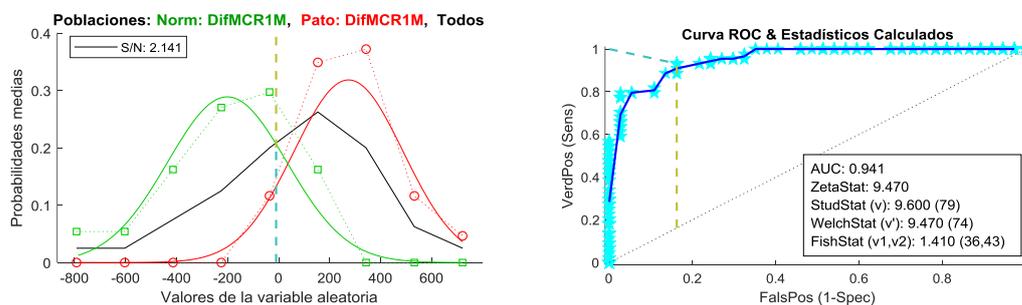


Figura 5.46 Histograma y curva ROC correspondientes al índice $I_{MCR1-MCR2}$.

El valor obtenido del estadístico F para las varianzas de este nuevo índice presentó un valor $F_{36,43}=1,410$ (Figura 5.46) que, por ser inferior a su correspondiente $F_{crítico}$ ($F_{36,43}^{0,01}=2,118$), permitió asumir la homogeneidad de las varianzas y evaluar la diferencia entre las medias poblacionales con el test de Student. El valor obtenido para el estadístico t ($t_{79} = 9,600$) fue mucho mayor al t-crítico correspondiente ($t_{79}^{0,01}=2,379$), por lo que pudo afirmarse que existe una diferencia significativa entre las medias de este índice para sanos y enfermos. El AUC calculado a partir de su curva ROC (AUC=0,941) indica, además, que este índice es el que mejor rendimiento presentó entre los testeados en esta sección.

5.7 Análisis Discriminante Lineal (LDA)

Tal como se explicó en el Capítulo 4, el Análisis Discriminante Lineal (LDA) es una técnica estadística multivariada cuya finalidad es analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos. Por lo tanto, con el objetivo de extraer de los espectros de fluorescencia la información más relevante para la clasificación de cada muestra según pertenezca a una persona sana o afectada por una patología oncológica, en la presente sección se describe la aplicación del LDA a diferentes matrices construidas a partir de los índices descriptos previamente.

Para realizar el análisis, la variable categórica de clasificación fue “*sanos*” o “*enfermos*”. El conjunto de espectros analizados fue el mismo con el que se trabajó a lo largo de todo el Capítulo, que corresponde a 81 muestras en total, (37 provenientes de personas “*sanas*” y 44 de “*enfermas*”), cada uno de los cuales registra 138 longitudes de onda de emisión. Dado que cada longitud de onda de emisión puede ser considerada una variable de análisis, y que esta cantidad era superior a la cantidad de objetos totales a analizar (81), fue necesario pre-procesar los datos de entrada al LDA, de manera de reducir su dimensionalidad. Cabe recordar que esto es condición necesaria para poder realizar el análisis, ya que uno de los supuestos que tiene que cumplir el LDA es, justamente, que el número de variables discriminantes tiene que ser menor que el número de objetos menos el número de categorías (en este caso igual a 2).

Por lo tanto, dado que no resulta conveniente aplicar esta técnica estadística a los espectros completos, se analizó cuál sería la mejor manera de reducir la dimensionalidad de los datos, o bien de generar espacios reducidos, donde se conserven ciertas características de los espectros que sirvan para luego separar apropiadamente ambas poblaciones. Esta es, justamente, la misma línea en la que se trabajó a lo largo de todo el presente Capítulo, habiendo logrado importantes resultados luego de analizar las más diversas metodologías.

En este contexto se buscó aprovechar los resultados previamente descriptos, que incluyen el análisis de cientos de índices, para potenciarlos con la capacidad discriminante del LDA. De esta manera se analizaron por LDA diferentes conjuntos de los mejores índices obtenidos en las secciones anteriores, mediante una rutina desarrollada especialmente a tal fin en el entorno Matlab.

De todo el análisis presentado en la sección 5.3, los mejores índices obtenidos fueron: $I_{471/507}$, $I_{471/540}$, $I_{471/575}$ e $I_{507/540}$, $I_{Area471}$, $I_{Area507}$, $I_{Area471-507}$ e I_{Posmax} , que se detallan en la tabla 5.3. Dado que algunos de los índices de este conjunto de datos poseen valores muy diferentes entre sí por ser de diferente origen (intensidad de fluorescencia a una determinada longitud de onda, cocientes y restas de éstas, e incluso un número que indica la posición del máximo), es necesario hacer algunas consideraciones antes de realizar el

análisis. En primer lugar es de interés que el algoritmo se ocupe de analizar la variabilidad entre los datos para cada variable, independientemente de su valor medio. Para lograr esto a cada valor se le restó el valor del promedio de esa variable, es decir se centraron los datos. Esto es, para una matriz **D** de datos, que consta de N filas en las que se ubican las muestras, y M columnas correspondientes a los 8 índices, la matriz de datos centrados **D_c** se calculó como:

$$D_c(i, j) = D(i, j) - \mu(j)$$

$$\mu(j) = \frac{1}{N} \sum_{i=1}^N D(i, j)$$

Otra cuestión importante para el análisis, es que los valores de todas las variables varíen dentro del mismo rango, es decir, que tengan valores similares entre sí. Si esta condición no se cumple el algoritmo le dará más peso a una variable sobre otra, solamente por tener valores de mayor magnitud. Para evitar este inconveniente generalmente se aplica una etapa de estandarización antes de realizar el análisis. El procedimiento de estandarización empleado en el presente trabajo de Tesis fue realizado aplicando las siguientes ecuaciones

$$D_{EC}(i, j) = \frac{D_c(i, j)}{Desv(j)}$$

$$Desv(j) = \sqrt{\frac{1}{N-1} \sum_{k=1}^N [D(k, j) - \mu(j)]^2}$$

donde **D_{EC}** es la matriz de datos **D** estandarizada y centrada.

Tanto la estandarización como el centrado se realizaron automáticamente con una rutina específica generada a tal fin en Matlab.

Para este primer grupo de análisis, se corrió la rutina LDA con la matriz de 81 muestras x 8 índices, previamente centrada y estandarizada. El resultado obtenido fue un vector, cuyos elementos (**I_{LDA-8}**) representan la proyección de cada muestra sobre el eje de Fisher, que es la línea que, de entre todas las que separan las clases, maximiza la separación entre ambas poblaciones sobre la base de los 8 índices extraídos de la sección 5.3. La figura 5.47 muestra las proyecciones obtenidas.

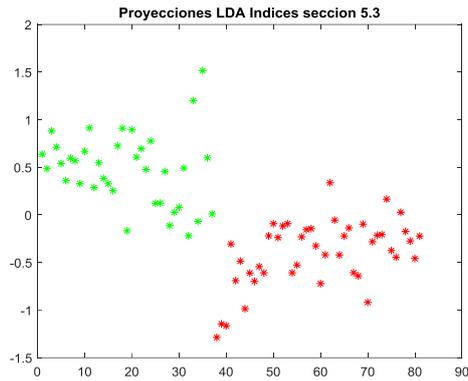


Figura 5.47 Proyecciones de I_{LDA-8} sobre el eje de Fisher

Para poder comparar los resultados obtenidos en cuanto al poder discriminante de este nuevo índice, se evaluaron las proyecciones obtenidas por LDA (I_{LDA-8}), haciendo uso de la rutina para análisis estadístico descripta en la sección 5.3. En la figura 5.48 se muestran los histogramas y la curva ROC correspondientes, junto con los valores estimados para los diferentes estadísticos.

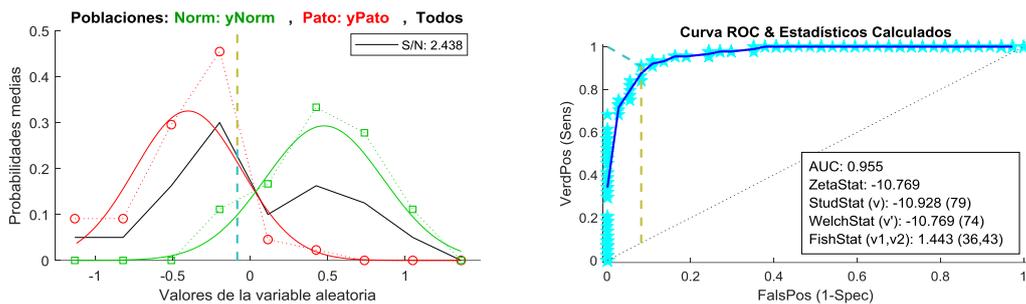


Figura 5.45 Histograma (Izq.) y curva ROC (Der.) correspondientes las proyecciones en el eje de Fisher de los índices de la sección 5.3.

Del análisis estadístico realizado a partir de los valores estimados que se muestran en la figura, se desprende que existe una diferencia significativa entre las medias de I_{LDA-8} para sanos y enfermos, por lo que es viable como índice de clasificación. Sin embargo, el resultado global de este nuevo test no es muy diferente del obtenido para el mejor de los índices de la sección 5.3. Es decir, el rendimiento del test, estimado a partir de su AUC es sólo ligeramente superior al obtenido con el índice $I_{471/507}$.

Otro conjunto de índices que se planteó, a partir de la necesidad de reducir la dimensionalidad de los datos de entrada para poder analizar por LDA, fue el de los scores del análisis de componentes principales (PCA). Recordando que, tal como se explicó en el Capítulo 4, el PCA es una herramienta que se utiliza para reducir la dimensionalidad de un conjunto de datos con el objetivo de mantener la mayor variabilidad de los datos

originales se empleó esta técnica como pretratamiento de los espectros antes de aplicar el LDA. Tal como se explicó anteriormente, el PCA descompone la matriz de datos original (\mathbf{X}), en dos matrices \mathbf{T} (scores) y \mathbf{P} (loadings) de menor dimensión. El análisis presentado en la sección 5.4 muestra que las primeras 8 componentes principales conservan el 99,82% de la variación total de la matriz de datos original. Entonces, se realizó el análisis por LDA a los primeros 8 scores, es decir se estudió mediante LDA a la matriz \mathbf{T} (81x8) obtenida por PCA, y luego se realizó el análisis estadístico de las proyecciones obtenidas. Cabe destacar que en este caso, debido a las condiciones impuestas, todas las variables que ingresan al LDA son de las mismas características, y sus valores numéricos se encuentran en la misma escala, por lo que no fue necesario el pre-procesamiento de los datos que implica el centrado y estandarizado de las variables.

Las proyecciones sobre el eje de Fisher obtenidas $I_{LDA-PCA}$ pueden observarse en la figura 5.46, mientras que la figura 5.47 muestra los histogramas y la curva ROC correspondientes.

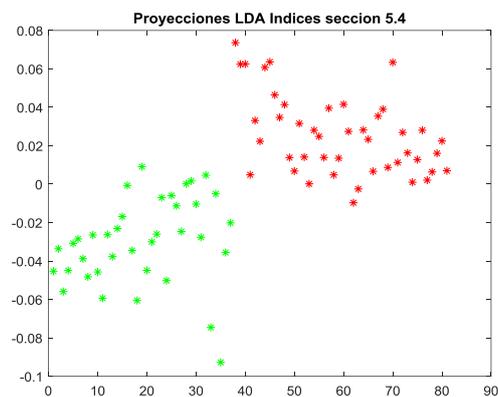


Figura 5.46 Proyecciones de $I_{LDA-PCA}$ sobre el eje de Fisher

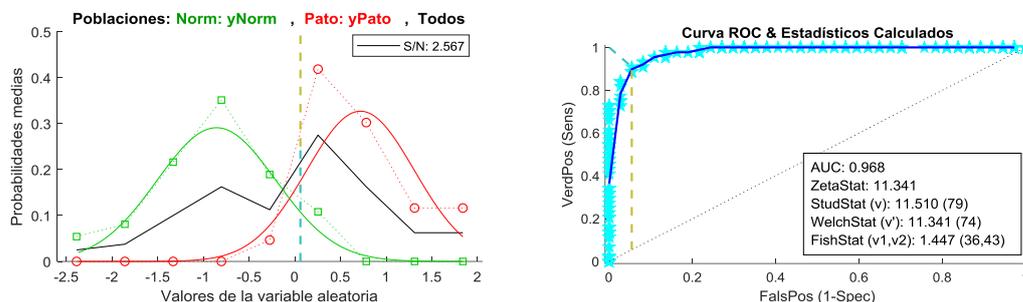


Figura 5.47 Histograma (Izq.) y curva ROC (Der.) correspondientes las proyecciones en el eje de Fisher de los índices de la sección 5.4.

Del mismo modo que en los casos anteriores, se realizó el análisis estadístico y se verificó la existencia de una diferencia significativa entre las medias de $I_{LDA-PCA}$ para sanos y enfermos. Finalmente, de la comparación de estos resultados con los obtenidos a partir de

los mismos datos en la sección 5.4, se concluye que en este caso la incorporación del LDA aumenta significativamente la eficiencia en la clasificación. Cabe recordar que del análisis estadístico realizado sobre el componente principal 1, en la sección 5.4, los valores obtenidos para la estimación de la relación señal/ruido y el AUC fueron 2,082 y 0,934, respectivamente, siendo el PC1 el que representa el 88,7% de la variabilidad total de los datos. Para el componente principal 2, que representa el 4,33% de la variabilidad, los valores obtenidos fueron 0,656 para la estimación de la relación señal/ruido y 0,661 para el AUC, mientras que para los componentes principales 3 en adelante estos valores fueron muy bajos, evidenciando su incapacidad para discriminar ambas poblaciones individualmente, y por ello se despreciaron. Por otra parte, en este nuevo análisis, al considerar los componentes principales 1 a 8 en conjunto y analizarlos por LDA, se obtuvo para $I_{LDA-PCA}$ una estimación de la relación señal/ruido 2,567 y un valor para AUC de 0,968, aumentando sustancialmente la potencia en la discriminación respecto de la de los PC individuales. En consecuencia, la utilización del PCA como herramienta de pre-tratamiento para reducir la dimensión de los datos de la matriz original (de 138 a 8, conservando el 99,82% de la variabilidad total), y luego el análisis por LDA se presenta, hasta aquí, como el test más eficiente en la discriminación de poblaciones.

El siguiente conjunto de datos que se analizó por LDA fue el formado por los mejores índices obtenidos en la sección 5.5. En dicha sección se analizó la utilización de índices generados a partir de los valores de intensidad de fluorescencia de emisión a cada longitud de onda, empleando tanto los espectros completos como sus derivadas primera y segunda. De todos los índices analizados, los que mostraron mejor poder discriminatorio entre poblaciones se listaron en las tablas 5.5 a 5.7, y son los que se utilizaron para realizar el análisis por LDA.

En la figura 5.48 se muestran los valores de las proyecciones en el eje de Fisher obtenidos para los índices de la sección 5.5 ($I_{LDA-S5.5}$), mientras que los correspondientes histogramas así como la curva ROC se muestran en la figura 5.49.

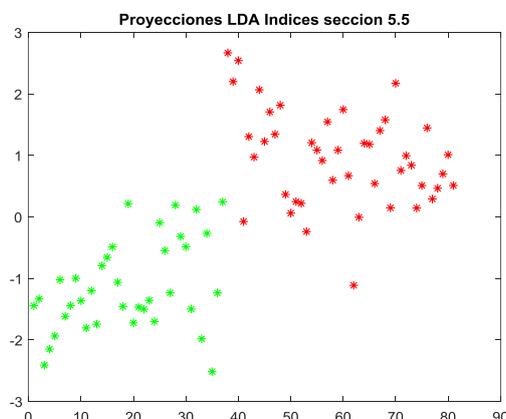


Figura 5.48 Proyecciones de $I_{LDA-S5.5}$ sobre el eje de Fisher

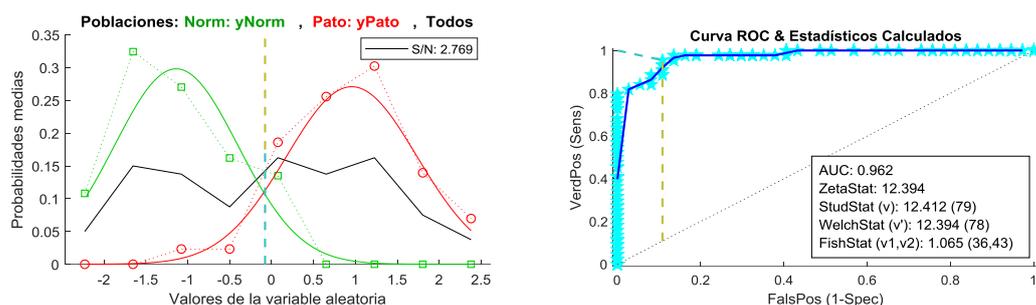


Figura 5.49 Histograma (Izq.) y curva ROC (Der.) correspondientes las proyecciones en el eje de Fisher de $I_{LDA-S5.5}$

Del análisis estadístico se evidencia que en este caso el análisis por LDA muestra resultados similares a los obtenidos en la sección 5.5 para algunos de los índices. Es decir, la eficiencia en la discriminación de poblaciones obtenida por LDA con el conjunto completo de índices generados en la sección 5.3 (AUC=0,962), es similar a la que presentan algunos índices individualmente ($I_{Eucl511}$: AUC=0,959, I_{D1-489} , I_{D1-551} e I_{D1-579} : AUC entre 0,961 y 0,963). Por lo tanto, si bien la potencia en clasificación obtenida mediante el análisis por LDA es comparable a las mayores obtenidas en el análisis realizado en la sección 5.5, el eje de Fisher no logra mejoras sustanciales en cuanto a la eficiencia del test, respecto de algunas variables.

Completando el análisis de todos los grupos de índices estudiados en el presente Capítulo, se realizó el análisis por LDA a los cuatro índices obtenidos en la sección 5.6. Dichos índices son los coeficientes (o scores) de cada una de las 4 contribuciones o componentes con las que se ajustó la matriz de los 81 espectros originales, mediante MCR-ALS. Cabe destacar, que el último caso estudiado al final de dicha sección, no se incluye explícitamente aquí por ser una combinación lineal (resta) de dos de los coeficientes en estudio, y realizar ese tipo de transformación es, justamente, el objetivo del LDA.

Las proyecciones sobre el eje de Fisher obtenidas (índice $I_{LDA-MCR}$) pueden observarse en la figura 5.50, mientras que la figura 5.51 muestra los histogramas y la curva ROC correspondientes.

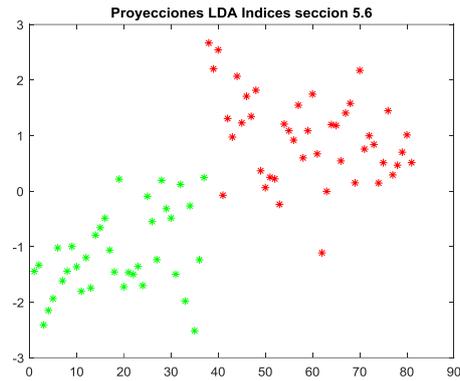


Figura 5.50 Proyecciones $I_{LDA-MCR}$ sobre el eje de Fisher

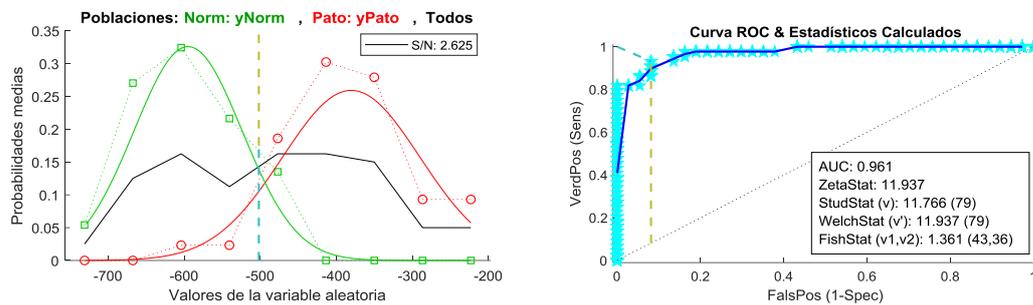


Figura 5.51 Histograma (Izq.) y curva ROC (Der.) correspondientes las proyecciones en el eje de Fisher de $I_{LDA-MCR}$

Los resultados indican un aumento en el poder de clasificación del índice $I_{LDA-MCR}$ respecto del obtenido para los cuatro índices analizados individualmente en la sección 5.6. Asimismo, cabe destacar que el índice obtenido a partir de la diferencia de los scores de los componentes 1 y 2, presentó un desempeño muy similar al obtenido aplicando LDA.

A partir de los resultados obtenidos en esta sección para cada uno de los grupos analizados en las secciones 5.3 a 5.6, y a modo de cierre, surgió la inquietud de comprobar si un análisis por LDA de todos los grupos de índices juntos podía mejorar realmente el desempeño del test. Es decir, si ingresando al LDA todos los índices que fueron considerados como buenos clasificadores poblacionales individuales a lo largo del presente Capítulo se podía encontrar un eje de Fisher en el que las proyecciones de ambas poblaciones se separen con mayor claridad.

La matriz de datos quedó constituida entonces por los 28 índices, o variables que se listan en la tabla 5.9.

Indice	Sección en la que fue analizado	Indice	Sección en la que fue analizado
$I_{471/507}$	Sección 5.3	I_{PC7}	Sección 5.4
$I_{471/540}$	Sección 5.3	I_{PC8}	Sección 5.4

$I_{471/575}$	Sección 5.3	$I_{Eucli465}$	Sección 5.5
$I_{507/540}$	Sección 5.3	$I_{Eucli511}$	Sección 5.5
$I_{Area471}$	Sección 5.3	$I_{Eucli597}$	Sección 5.5
$I_{Area507}$	Sección 5.3	I_{D1-485}	Sección 5.5
$I_{Area471-507}$	Sección 5.3	I_{D1-489}	Sección 5.5
I_{PosMax}	Sección 5.3	I_{D1-551}	Sección 5.5
I_{PC1}	Sección 5.4	I_{D1-579}	Sección 5.5
I_{PC2}	Sección 5.4	I_{D2-509}	Sección 5.5
I_{PC3}	Sección 5.4	I_{MCR1}	Sección 5.6
I_{PC4}	Sección 5.4	I_{MCR2}	Sección 5.6
I_{PC5}	Sección 5.4	I_{MCR3}	Sección 5.6
I_{PC6}	Sección 5.4	I_{MCR4}	Sección 5.6

Tabla 5.9.

Dado que la mayoría de los índices poseen valores de diferentes magnitudes, debido a sus múltiples orígenes, en este caso fue necesario centrar y estandarizar todas las variables antes de aplicar el análisis. Luego, con la rutina generada en Matlab se realizó el análisis por LDA, obteniéndose las proyecciones sobre el eje de Fisher (I_{LDA-28}) que se muestran en la Figura 5.52.

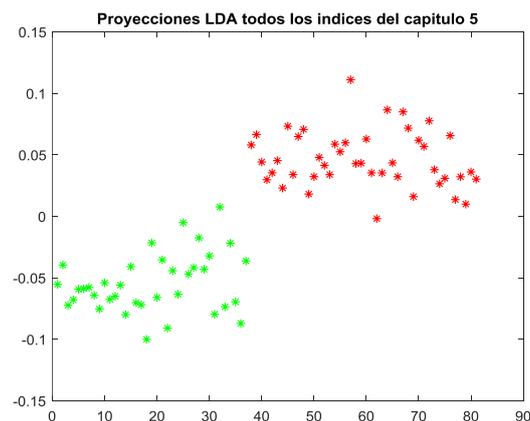


Figura 5.52 Proyecciones de I_{LDA-28} sobre el eje de Fisher

A partir de la observación de la figura, se podría afirmar que el nuevo índice I_{LDA-28} tiene la capacidad de discriminar muy bien entre ambas poblaciones, es decir, se observa que las medias muestrales para sanos y enfermos son claramente diferentes. No obstante

ello, se verificó numéricamente si dicha diferencia era estadísticamente significativa mediante el análisis estadístico descrito previamente. La tabla 5.8 muestra los valores obtenidos para los diferentes parámetros estadísticos.

Índice	S/N	F-Fisher	t-Student	t-Welch (f')	AUC
I_{LDA-28}	1,816	1,304 (36,43)	19,632	19,431 (75)	0.986

Tabla 5.8 Resultados del análisis estadístico de I_{LDA-28}

Si bien dada la cantidad de muestras de cada grupo se podría asumir la condición de normalidad, se analizaron los histogramas que se presentan en la figura 5.53 para comprobar esta hipótesis.

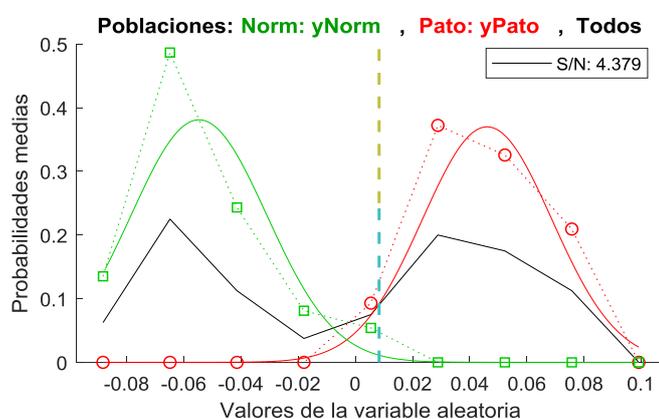


Figura 5.53 Histogramas correspondientes las proyecciones en el eje de Fisher de los índices de la tabla 5.7

La comparación de los histogramas con las campanas de Gauss estimadas para cada población a partir de los correspondientes parámetros muestrales permitió verificar explícitamente que las proyecciones de cada grupo sobre el eje de Fisher presentan una distribución aproximadamente normal. Posteriormente se realizó la prueba F para analizar la homogeneidad de las varianzas. El valor del estadístico F para las varianzas obtenido ($F_{36,43}=1,304$) fue menor que el $F_{crítico}$ ($F_{36,43}^{0,01}=2,16$) correspondiente, por lo que se asumió la homogeneidad de las varianzas. Entonces, se evaluó la diferencia entre las medias poblacionales con el test de Student. El valor t obtenido $t=19,632$ fue mucho mayor que el valor $t_{crítico}$ ($t_{79}^{0,01}=2,379$), por lo que se concluye que existe una diferencia significativa entre las medias de este índice para sanos y enfermos ($p<0,0005$).

A partir de estos resultados se puede afirmar que el índice formado por los valores de las proyecciones sobre el eje de Fisher para el conjunto completo de datos de la tabla 5.7 (I_{LDA-28}), tiene la capacidad de clasificar a las muestras de orina, según sean de personas

sanas o enfermas, ya que corresponden a distribuciones de probabilidad con medias poblacionales claramente distintas.

Con el fin de evaluar la capacidad del índice para el desarrollo de un test que permita la discriminación de ambas poblaciones se realizó la curva ROC correspondiente. Tal como puede observarse en la figura 5.54 la curva ROC obtenida presentó una gran similitud con la curva ROC “ideal”, o de clasificación perfecta, ya roza la coordenada (0,1) del espacio ROC. Recordemos que esa coordenada representa un 100% de sensibilidad (ningún falso negativo) y un 100% también de especificidad (ningún falso positivo). El alto valor de AUC obtenido (0,986) indica un excelente rendimiento diagnóstico, ya que existe un 98,6% de probabilidad de que ante un par de individuos, uno enfermo y el otro sano, la prueba los clasifique correctamente.

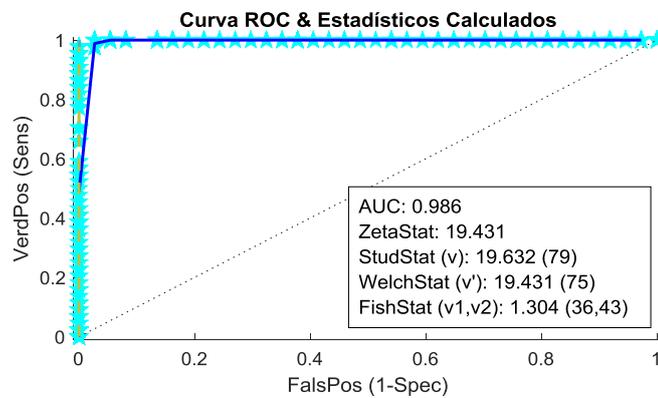


Figura 5.54 Curva ROC correspondientes las proyecciones en el eje de Fisher de los índices de la tabla 5.7 (I_{LDA-28})

Del análisis expuesto en el presente Capítulo, se concluye que de todos los índices testeados, el que mostró la mejor capacidad para discriminar entre ambas poblaciones es el obtenido por LDA a partir de todos los valores listados en la tabla 5.7 juntos (I_{LDA-28}).

Capítulo 6. Matrices de fluorescencia de muestras de Orina

El estudio de la topografía de las EEMs de la orina humana es una nueva variante en bioquímica analítica para el reconocimiento de patrones. Sin embargo la aplicación práctica de esta metodología requiere de la interpretación de las características espectrales y esto resulta complicado debido tanto a la variabilidad de las muestras como a la complejidad de las señales. En este sentido los topogramas urinarios (EEMs de las muestras de orina) varían mucho de muestra a muestra (o para una misma muestra con diferentes tiempos de conservación) tanto desde un punto de vista cuantitativo como cualitativo. [81] Dado que esto puede constituir una gran limitación a la hora de interpretar los resultados, se ha tenido especial cuidado en la dilución de las muestras. Asimismo, debido a que el pre tratamiento de los datos de emisión resulta crítico para el análisis multivariado, se aplicaron diferentes estrategias para la reducción de efectos no lineales.

6.1 Análisis multivariado de la fluorescencia de orina

Tal como se ha mencionado la sección 1.3.5, el análisis de fluidos corporales es una gran fuente de información para el diagnóstico médico. En la actualidad la mayoría de los pacientes son diagnosticados gracias a los datos de estos análisis, siendo los más utilizados los análisis de sangre y de orina.

En particular, la orina es un fluido biológico con múltiples componentes cuya concentración puede variar fisiológicamente debido a efectos tales como la edad del paciente, el ritmo circadiano, el estado nutricional, el grado de hidratación y la actividad física. Sin embargo, a pesar de la complejidad composicional de la orina, en general se pueden observar entre 3 y 5 máximos de fluorescencia. Los máximos observados no resultan de un único metabolito fluorescente sino que son el resultado de varias especies que poseen rasgos espectroscópicos similares, que en muchos casos son consecuencia directa de sus similitudes estructurales. Para estudiar las características de la emisión fluorescente de la orina en un amplio rango se recurre al estudio de sus denominadas matrices de excitación/emisión (MEE). En el presente Capítulo se analizó la capacidad diagnóstica de las MEE las muestras de orina, mediante el análisis paralelo de factores (PARAFAC).

Análisis paralelo de factores.

El análisis paralelo de factores (PARAFAC) ha sido presentado en detalle en la sección 4.3.4. Es una generalización del análisis por componentes principales (PCA) que se aplica frecuentemente para el análisis de datos espectrofluorométricos, en particular a conjuntos de matrices de excitación-emisión (MEE) ordenados en arreglos de 3 vías (**X**).

Dada una estructura \underline{X} , que contiene las MEE registradas para varias muestras, el algoritmo PARAFAC descompone a \underline{X} en tres matrices **A**, **B** y **C**, que contienen, los perfiles relativos de concentración, los espectros de emisión y los espectros de excitación correspondientes a cada uno de los componentes presentes en las muestras analizadas. Aunque el algoritmo es más complejo que otros empleados para descomposiciones lineales, la solución también se obtiene mediante optimización por mínimos cuadrados alternados y durante el proceso iterativo se pueden imponer restricciones a las soluciones para que éstas resulten más fáciles de interpretar o se ajusten mejor a las características físicas del sistema estudiado. Cabe destacar que frecuentemente los arreglos de datos obtenidos mediante la medición de MEE para diferentes composiciones obedecen a una estructura trilineal. En estos casos, si el número de factores se elige correctamente, PARAFAC es capaz de extraer sin ambigüedades tanto los perfiles de concentración como los perfiles espectrales (de excitación y de emisión), aún para sistemas con alto grado de solapamiento. Sin embargo, la construcción de un modelo PARAFAC para un conjunto de EEM no es algo trivial e involucra una serie de etapas que incluyen: el pre-procesamiento de las señales registradas, la exploración del conjunto de datos, el desarrollo del modelo y la validación del mismo. A continuación se detallan los pasos seguidos en cada una de las etapas para un conjunto de 100 MEE correspondientes a 52 muestras de orina de personas diagnosticadas con patologías oncológicas y 48 muestras de personas sanas.

Pre-procesamiento

Dado que el modelo PARAFAC se basa en la descomposición trilineal de las intensidades de emisión registradas en las MEE para diferentes muestras, resultó necesario el pre-procesamiento de los datos medidos experimentalmente. El pre-procesamiento incluyó las siguientes etapas:

Corrección de los efectos de filtro interno (IFE). Los IFE se corrigieron mediante el empleo de los espectros de absorción correspondientes a cada muestra, tal como fue explicado en la sección 4.3.4.

Corrección por el blanco, a partir de la sustracción de la matriz correspondiente a las emisiones del agua bidestilada. A cada MEE de orina se le restó la matriz del agua bidestilada medida ese mismo día.

Eliminación de las regiones de las MEE gobernadas por efectos de dispersión de la radiación. Se eliminaron las regiones o “calles” correspondientes a las dispersiones Rayleigh y la de Raman de primer y segundo orden. Se trabajó, entonces, con arreglos de 3 vías con elementos faltantes, que en el entorno de Matlab se indican como NaN (Not a Number).

Normalización. El conjunto de datos formado por las MEE de las muestras de orina puede contener matrices con valores de intensidad muy dispares entre sí. Esto es principalmente

debido a la variabilidad en la concentración de los metabolitos en las muestras. En este contexto resulta conveniente igualar el peso de todas las matrices ante el algoritmo PARAFAC para evitar que las muestras con mayor señal tengan más importancia en el ajuste del modelo que aquellas con baja señal. En consecuencia, se normalizó a cada muestra por su área bajo la curva (o intensidad de señal total), permitiendo al modelo centrarse en las variaciones cualitativas entre muestras en lugar de centrarse en la magnitud total de la señal.

Fase exploratoria

Esta etapa involucró dos objetivos principales: establecer el mejor conjunto de datos posible para el modelado y obtener una idea preliminar del número correcto de factores para el modelo. Para lograr el primero de los objetivos es necesario identificar los denominados “outliers”. Estos outliers pueden ser muestras o longitudes de onda (tanto de emisión como de excitación), poco representativos o de baja calidad. En cada fase exploratoria realizada a lo largo del presente trabajo de Tesis, los criterios para la identificación y eliminación de outliers fueron:

- Evaluación del rango óptimo: se eliminaron de los datos de las MEE correspondientes a longitudes de onda de excitación que estuvieran por debajo de 275 nm. Esto es debido a que en esa zona, los espectrofluorómetros utilizados no presentan buen rendimiento, ya que poseen baja emisión de la lámpara y baja sensibilidad de los detectores, lo que deriva en una mala relación señal/ruido.
- Identificación de muestras con baja correlación entre su espectro de absorción y su fluorescencia total. Para facilitar la identificación de estos outliers, se incorporó una rutina para creación automática de un gráfico de dispersión como el de la figura 6.1. Allí se representa cada muestra como un punto en el espacio (X,Y), donde el eje X corresponde a la estimación de la concentración total de analitos calculada como el área total bajo la curva del espectro de absorción y el eje Y a la estimación del total de las cuentas calculadas por el modelo PARAFAC para la muestra considerada. En esta representación las muestras que se apartan sustancialmente de la diagonal que correlaciona la emisión con la absorción son identificados como posibles outliers. En la figura 6.1, a modo de ejemplo, se señalan los puntos correspondientes a las muestras 005 y 089 como posibles outliers. Luego, a partir del análisis detallado de los espectros de absorción de las muestras identificadas se resolvió su eliminación o no del conjunto de MEE.

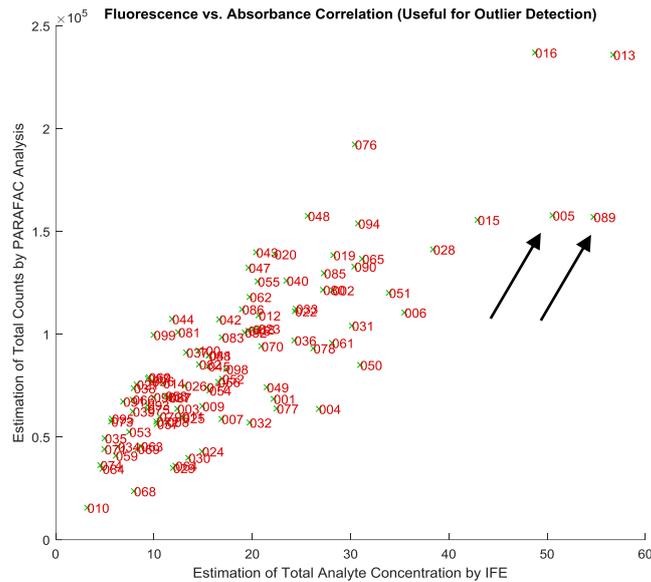


Figura 6.1 Ventana del software desarrollado que muestra la estimación total de cuentas calculadas por PARAFAC en función de la estimación de la concentración total de analitos

- Identificación de muestras que presenten altos valores de falta de ajuste de los datos. Dicha falta de ajuste, llamada LOF (del inglés Lack of fit), se calcula como la diferencia entre los datos de entrada y los datos reproducidos a partir del modelo.
- Identificación de muestras cuyos residuos presenten estructura o no se distribuyan aleatoriamente. Esta identificación se logró a partir de la inspección de las matrices de los residuos.

Una vez identificados, los outliers se examinaron individualmente para determinar el motivo más probable de su comportamiento atípico pudiendo ser eliminados del conjunto de datos. Para facilitar la inspección de la matriz de residuos, de las correcciones aplicadas, y de la contribución de cada factor a las muestras de las MEE estudiadas en el presente trabajo de Tesis, se desarrolló una rutina específica en el entorno de programación Matlab. Las ventanas del software desarrollado, diseñadas para analizar cada muestra individualmente, se muestran en las figuras 6.2 y 6.3.

La figura 6.2 muestra la MEE medida experimentalmente seguida de la matriz calculada por PARAFAC y finalmente la matriz de los residuos. Esta última es calculada a partir de la diferencia de las dos primeras y se muestra en una escala diferente de intensidades para facilitar el análisis de los residuos. Además, sobre el tercer recuadro se muestra el valor de la falta de ajuste porcentual (LOF) para esa muestra particular. Así esta ventana constituye una herramienta fundamental a la hora de analizar los residuos de cada muestra, ya que además de dar el valor de LOF para cada muestra, permite

inspeccionar su matriz de los residuos y verificar que no presenten ninguna estructura definida.

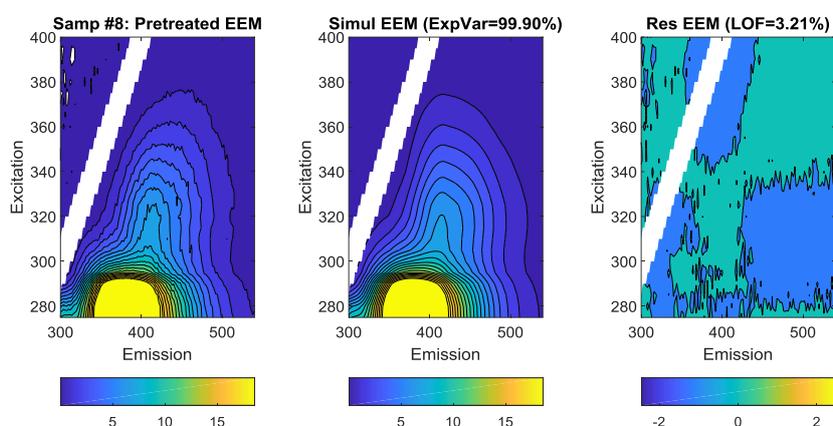


Figura 6.2: Ventana del software desarrollado que muestra la MEE medida experimentalmente (Izq.), la matriz calculada por PARAFAC (Centro), y la matriz de los residuos (Der.) para una muestra de orina.

La figura 6.3 muestra la ventana del software desarrollado que permite el análisis de los factores individuales. El recuadro inferior derecho muestra la MEE que se midió experimentalmente y el recuadro superior izquierdo la contribución del factor en estudio a la MEE a la muestra analizada (por ejemplo, la figura muestra la contribución del factor 2 a la muestra 14). Los otros recuadros muestran los espectros normalizados de emisión, excitación y absorción para esa muestra. La utilidad de esta ventana radica en que permite saber cuánto contribuye cada factor sobre la matriz medida experimentalmente. Para ejemplificar este punto, en la figura 6.4 se muestra la misma ventana de análisis para esa misma muestra pero con otro factor seleccionado (factor 4). Allí se puede observar que este último tiene una contribución mucho menor a la del factor 2 para esa muestra.

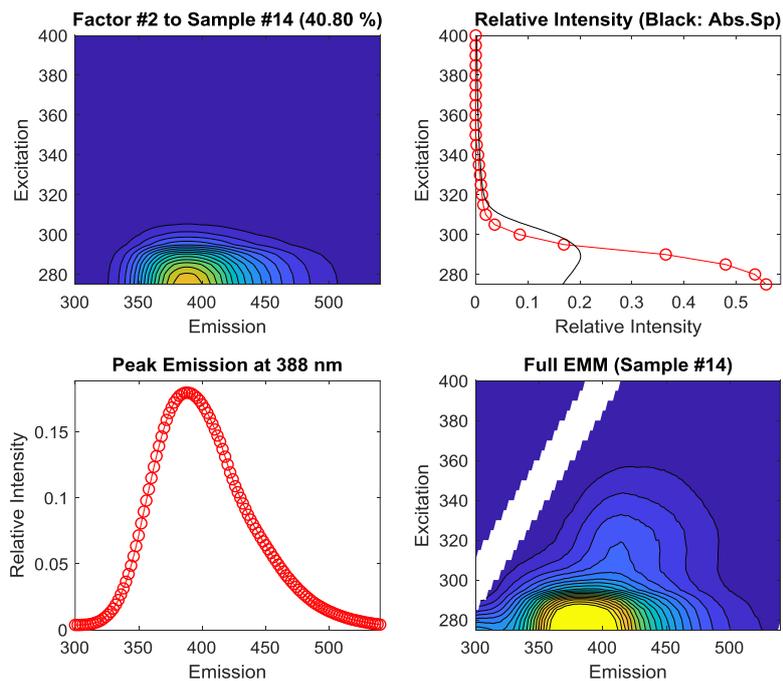


Figura 6.3: Ventana del software desarrollado que permite el análisis de los factores individuales.

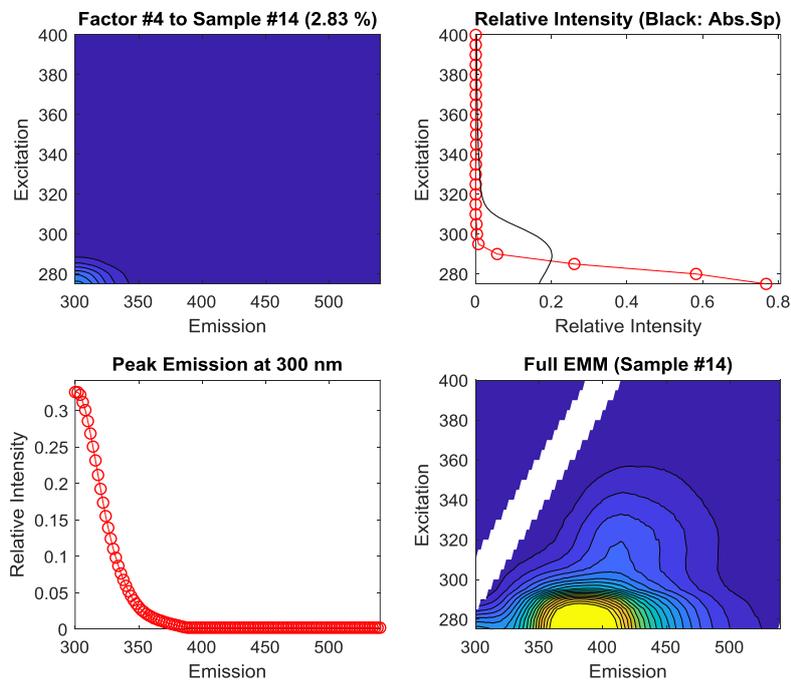


Figura 6.4 Ventana del software desarrollado que muestra la contribución del factor 4 a la muestra número 14.

Tal como se expuso previamente, la secuencia de pasos completa es fundamental para que el ajuste sea satisfactorio, y en general es necesario ejecutarla varias veces.

Por último, para determinar si en el modelo obtenido mediante PARAFAC se ha elegido el número correcto de componentes (factores), se consideraron varios criterios en forma simultánea para la evaluación de la calidad del modelo. En este contexto, para validar el modelo se tuvo en cuenta la falta de ajuste, la aleatoriedad de las matrices de los residuos, el análisis de la verosimilitud de los perfiles espectrales tanto de excitación como de emisión para cada factor y el diagnóstico de consistencia del núcleo (Core Consistency).

El análisis de la verosimilitud de los perfiles espectrales de los componentes implica valorar si los resultados del modelo PARAFAC son físicamente razonables con respecto al fenómeno en estudio. Dado que se trabajó con datos provenientes de MEE, se tuvieron en cuenta las características fundamentales de la espectroscopia de fluorescencia. Entonces a todos los análisis realizados por PARAFAC se aplicaron las siguientes restricciones: las matrices de los scores (**A**), espectros de excitación (**B**) y de emisión (**C**) sólo pueden poseer números positivos (restricción de no negatividad); los espectros de emisión, además deben presentar una única banda (unimodalidad). Luego, se verificó que además de cumplir esas restricciones, los espectros de excitación y de emisión de cada factor estén desplazados entre sí (emisión hacia mayores longitudes de onda), y presenten un mínimo de superposición entre ellos. Cumplidas estas condiciones, se consideró que los espectros obtenidos con el algoritmo eran físicamente aceptables. Como criterio complementario se aplicó el test de evolución de la consistencia del modelo, también llamado consistencia del núcleo (core consistency). Este test da una idea de si se cumple la condición de trilinealidad del modelo PARAFAC en relación a la cantidad de factores elegida. Si se ejecuta una secuencia de modelos con un número creciente de factores, el valor de este parámetro tiende a caer bruscamente cuando se emplean más factores que los necesarios.

6.2 Modelo PARAFAC de las MEE medidas

El análisis realizado en el Capítulo 5 mostró que a partir de la información contenida en los EEm de muestras de orina se pudo diseñar un test que permitió una buena discriminación entre personas sanas y personas afectadas por enfermedades oncológicas. Cabe recordar que a lo largo de ese Capítulo los espectros de emisión fueron obtenidos excitando con radiación de 405 nm. Por otra parte, es sabido que la mayoría de los analitos fluorescentes presentes en la orina emiten en la región espectral UV [82] y entonces, la mayor contribución a la fluorescencia total de la orina se da en esa zona.

Con el objetivo de capturar a la mayor cantidad de los fluoróforos presentes en la orina y profundizar el análisis realizado sobre los espectros, se extendió el rango espectral de estudio incorporando MEE que fueron registradas empleando un amplio rango de

longitudes de onda de excitación centrado alrededor de los 405 nm. Así se registraron MEE excitando con radiación de entre 250 y 500 nm, y barriendo las intensidades de emisión entre 275 y 700 nm. La figura 6.5 muestra una matriz representativa, tal como fue registrada (es decir, sin pre-procesamiento).

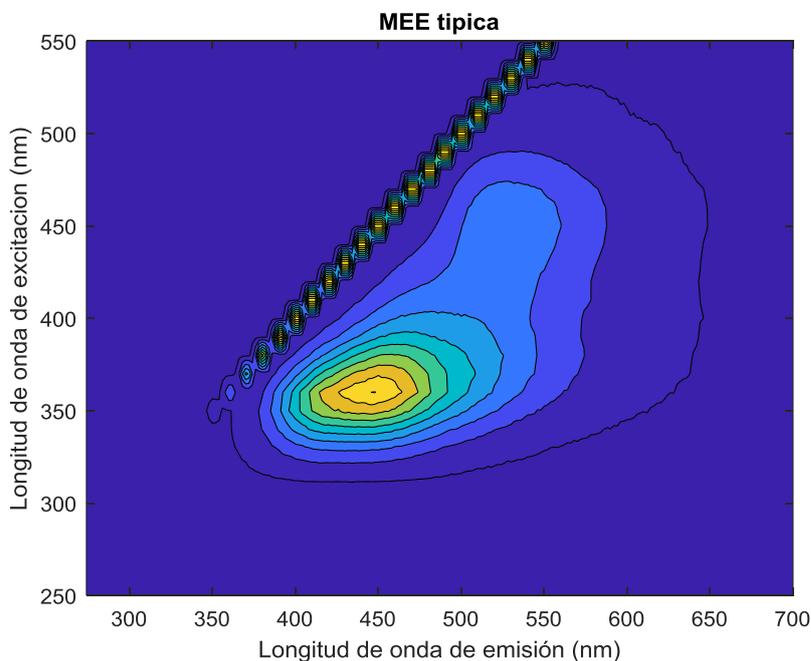


Figura 6.5 MEE típica, sin ningún pre-procesamiento.

Allí se observa que la MEE presenta un máximo de intensidad bien definido aproximadamente en Ex360/Em450. Sin embargo, en la región de longitudes de onda por debajo de 380 nm, donde era de esperarse la mayor intensidad de fluorescencia, la emisión fue muy pobre y prácticamente no se registró señal. En consecuencia, para analizar posibles causas de este comportamiento imprevisto se registraron y estudiaron los espectros de absorción de las muestras. En todos los casos se observó que éstos presentaban valores de absorbancia muy elevados a longitudes de onda por debajo de los 400 nm (figura 6.6). En consecuencia, la baja intensidad de la señal de fluorescencia registrada en la zona UV se asoció directamente a los IFE de las muestras de orina sin diluir. [83]

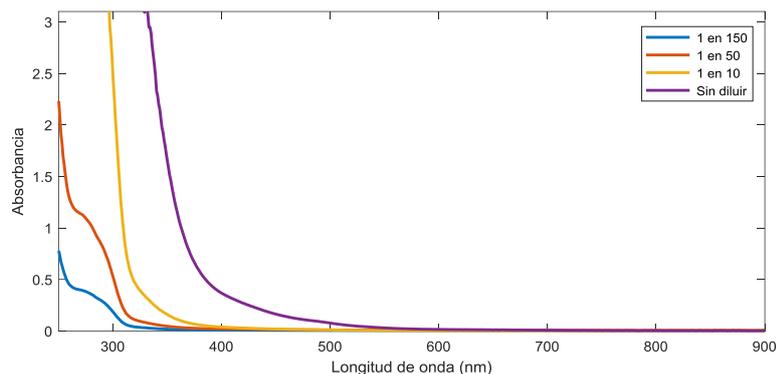


Figura 6.6. Espectros de absorción típicos de las muestras de orina sin diluir y con diluciones de hasta 1 en 150

Esto planteó la necesidad de pre-tratar las muestras, teniendo que diluirlas entre 150 y 700 veces antes de medir sus MEE con el fin de que la absorbancia máxima no fuera superior a 1,8 en todo el rango espectral analizado. De esta manera, la dilución de las mismas cumplió un rol fundamental a fin de garantizar que se trabaje dentro del rango de linealidad. Es importante mencionar además que la dilución se realizó con el buffer fosfato (previamente descrito en el Capítulo 4), para regular el pH y de esta manera estabilizar el comportamiento de los fluoróforos.

A partir de aquí se presenta, entonces, el análisis de un conjunto de datos formado por las MEE de 100 muestras de orina registradas en el rango de longitudes de onda de excitación de 260 a 400 nm (cada 5 nm), y de emisión de 300 a 600 nm (cada 2 nm). De esta manera el conjunto de datos a analizar quedó conformado por 100 MEE de dimensión 29 x 131 cada una, 48 de las cuales correspondían a la fluorescencia de la orina de personas sanas y 52 a la de personas diagnosticadas con enfermedades oncológicas. Todas las muestras fueron pre-procesadas según los procedimientos descritos en la sección 6.1, con la consideración especial en cuanto a la dilución. A partir de la evaluación del rango óptimo se descartaron los datos de las MEE correspondientes a zonas de excitación con longitudes de onda por debajo de 275 nm y a zonas de emisión con longitudes de onda de superiores a 540 nm. Las primeras longitudes de onda fueron descartadas por cuestiones instrumentales del equipo y las últimas porque se observó que no aportaban información al análisis. Luego, con el objetivo de identificar las muestras que pudieran ser clasificadas como outliers, se analizaron sus respectivos espectros de absorción. Este análisis constituyó un paso fundamental del pre-procesamiento de los datos, y por eso se describirá aquí en detalle. Su importancia radica en que un espectro de absorción defectuoso o mal medido puede derivar en malas correcciones en la MEE de la muestra correspondiente y condicionar sustancialmente la calidad de ajuste del modelo. En una

primera etapa se identificaron las muestras que presentaban una mala correlación entre la absorbancia y la fluorescencia total en el gráfico generado a tal fin (figura 6.7).

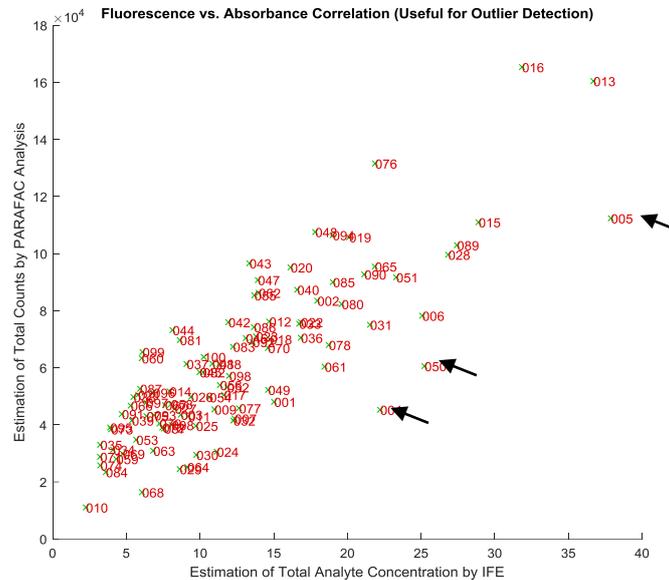


Figura 6.7 Posibles outliers señalados en el gráfico de correlación de absorbancia y fluorescencia total

A partir de esta figura surgieron como posibles outliers las muestras indicadas con flechas negras, que corresponden a las muestras 004, 005 y 050, que presentan un apartamiento apreciable respecto de la correlación general. A continuación se estudiaron en detalle los espectros de absorción de todas las muestras, y en particular de las tres mencionados previamente. En la figura 6.8 se muestran todos los espectros.

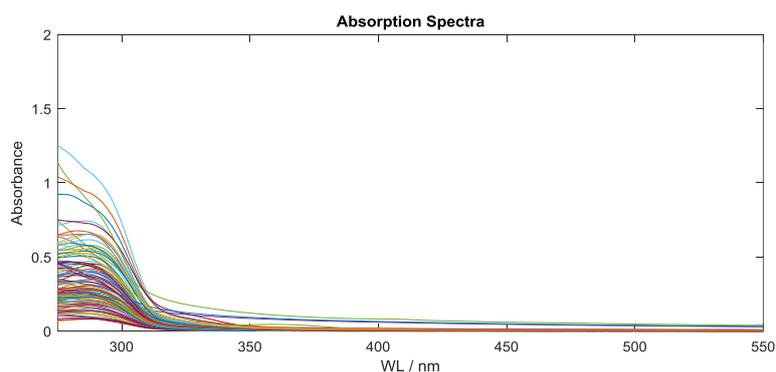


Figura 6.8: Espectros de absorción de las 100 muestras estudiadas.

Si bien la mayoría de las curvas parecen corresponder con espectros de absorción razonables, algunas de ellas presentan valores de absorbancia considerables para longitudes de onda superiores a 350 nm. Para apreciar mejor ese punto se incluye la figura

6.9, en la que se muestran los mismos espectros de absorción de la figura 6.8 pero con la escala de las ordenadas expandida.

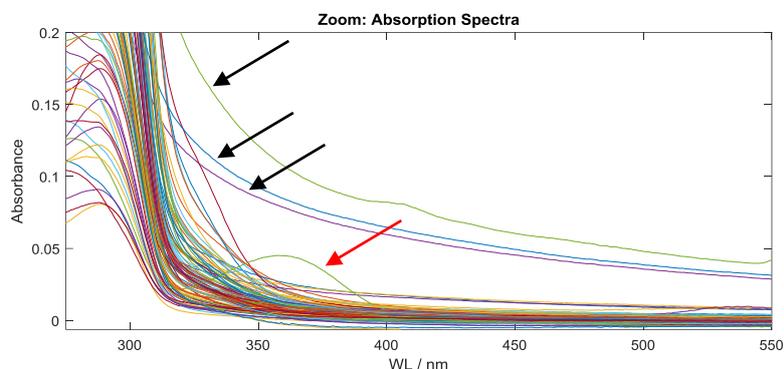


Figura 6.9. Zoom de la figura 6.8 donde se indican con flechas negras los espectros correspondientes a las muestras 004(violeta), 005 (verde) y 050 (celeste). La flecha roja indica el espectro de la muestra 068.

Puede apreciarse que los espectros indicados con flechas negras presentan su línea base desplazada muy por encima del error instrumental. Dichos espectros son los correspondientes a las muestras 004 (en violeta), 005 (en verde) y 050 (en celeste), que justamente son mismas que se señalaron como sospechosas en el gráfico de correlación entre absorción y fluorescencia total (figura 6.7). Entonces, dado que estos espectros de absorción son inaceptables, las muestras correspondientes fueron consideradas outliers y se eliminaron del análisis. Por otra parte, la flecha roja de la figura 6.9 señala el espectro de la muestra 068. Este espectro también es considerado como posible outlier por presentar una amplia banda de absorción centrada en 360 nm que lo diferencia del resto, por lo que fue analizado en detalle en un paso posterior. Esta banda podría ser indicio de contaminación de la muestra o bien de algún estado particular de ese paciente (metabólico o patológico). Por último, las muestras 013, 016 y 089 (celeste, verde y naranja en la figura 6.6) presentaron valores de absorbancia mayores a 1 en longitudes de onda por debajo de los 300 nm por lo que también fueron posteriormente estudiadas en detalle.

Finalmente, cabe destacar que el análisis de los espectros de absorción revela que ninguna muestra presentó valores de absorbancia mayores a 1,8 en todo el rango de interés (figura 6.8), indicando que las diluciones realizadas fueron apropiadas.

6.2.1 Determinación de la cantidad de factores

Tal como se explicó previamente, establecer el número óptimo de factores para el modelo es parte del “arte” del PARAFAC: no existe una receta infalible y siempre implica considerar algunas relaciones de compromiso entre diferentes criterios. La estrategia

habitualmente empleada para determinar el número correcto de factores consiste en ejecutar una secuencia de modelos PARAFAC incrementando el número de factores y comparar los resultados obtenidos sobre la base de los criterios de validación descritos al final de la sección 6.1. En el presente trabajo de Tesis se compararon modelos de 2 a 5 factores. A continuación se presenta el análisis comparativo de los modelos generados por el algoritmo. Los espectros de excitación y de emisión obtenidos se muestran en las figuras 6.10 a 6.13. En todos los gráficos los factores se han ordenado según su energía de emisión ascendente. Es decir que de entre todos los factores, el número 1 es aquél cuya banda de emisión se encuentra a la mayor longitud de onda, el número 2 el que le sigue, y así sucesivamente.

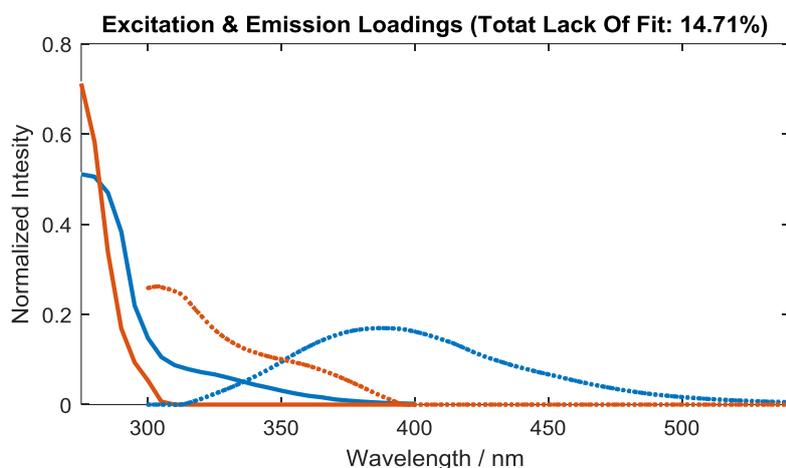


Figura 6.10 Perfiles de excitación y emisión del Modelo PARAFAC con 2 factores

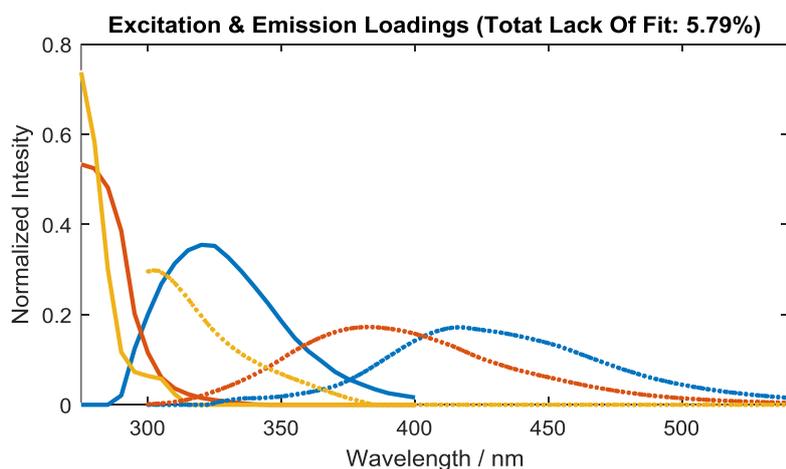


Figura 6.11 Perfiles de excitación y emisión del Modelo PARAFAC con 3 factores

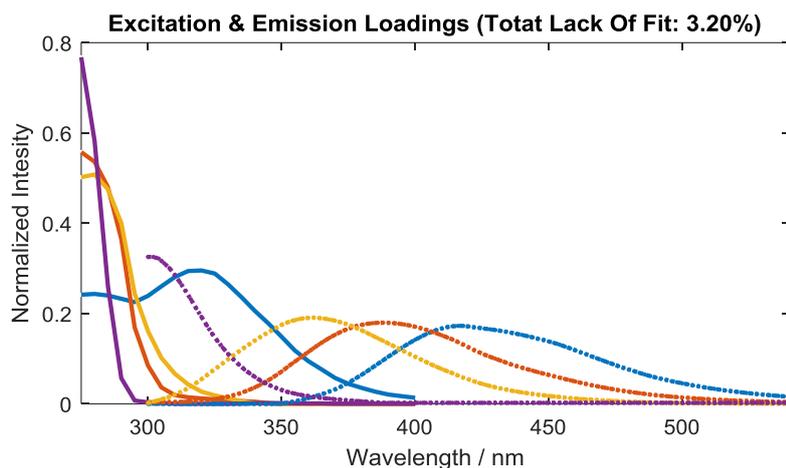


Figura 6.12 Perfiles de excitación y emisión del Modelo PARAFAC con 4 factores

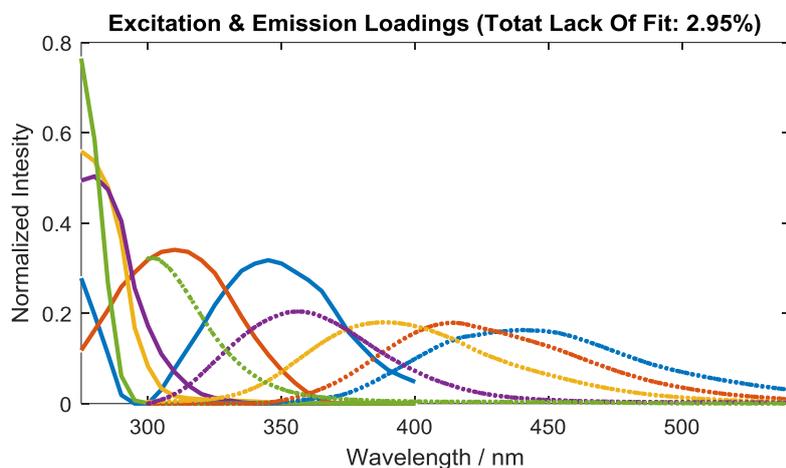


Figura 6.13. Perfiles de excitación y emisión del Modelo PARAFAC con 5 factores.

Para tener una idea de la bondad del ajuste del modelo PARAFAC en relación a la cantidad de factores elegida se calcularon los valores del test de evolución de la consistencia del modelo. En la tabla 6.1 se listan dichos valores junto con el porcentaje estimado para la falta de ajuste global (LOF total) asociada a cada modelo.

Cantidad de Factores del modelo PARAFAC	Consistencia del núcleo	Falta de ajuste global (LOF)
2	84,6%	14,7%
3	95,63%	5,79%
4	75,98%	3,20%
5	7,03%	2,95%

Tabla 6.1 Valores obtenidos del test de evolución de la consistencia del modelo y porcentaje estimado de la falta de ajuste global para los modelos PARAFAC con 2, 3, 4 y 5 Factores

En general, el valor de la consistencia del núcleo decrece a medida que aumenta el número de factores. Además el valor de este parámetro tiende a caer bruscamente cuando se emplean más factores que los necesarios. Este último punto se evidencia claramente en la tabla 6.1, donde se observa que al aumentar el número de factores de 4 a 5, la consistencia del núcleo cayó de un 75,98% a un 7,03%. Asimismo, la disminución de la falta de ajuste al pasar de 4 a 5 factores es muy poco significativa. En consecuencia, tanto la consistencia del núcleo como la falta de ajuste global sugieren que el modelo de 5 factores puede presentar problemas de sobreparametrización. Finalmente el análisis de la figura 6.13 revela que el espectro de excitación obtenido para el factor 1 del modelo con 5 factores (curva azul) es inaceptable físicamente, ya que posee una brusca caída a cero alrededor de los 300 nm, seguida de un importante incremento a longitudes de onda más largas. Por lo tanto, sobre la base de la información presentada en la tabla 6.1 y en la figura 6.13, el modelo de 5 factores quedó completamente descartado.

Por otro lado, tal como era de esperarse, el porcentaje de falta de ajuste (LOF) varía en sentido contrario a la consistencia del núcleo: disminuye a medida que se incrementa el número de factores del modelo. En esta secuencia de modelos la falta de ajuste para 2 factores fue realmente mala, ya que presentó un valor de 14.7% (tabla 6.1). Para el modelo con 3 factores decreció a valores aceptables (5,79%) y con 4 mejoró aún más (3,20%). Tomando en consideración tanto los altos valores de LOF global como el análisis de la estructura de las matrices de los residuos correspondientes, el modelo de 2 factores también fue descartado.

Los modelos PARAFAC con 3 y con 4 factores presentaron valores razonables tanto para el LOF como para la consistencia del núcleo. Entonces, con el fin de seleccionar el número óptimo de factores para el modelo, se recurrió al análisis detallado de la forma de sus espectros de excitación y de emisión. Las figuras 6.14 y 6.15 muestran los espectros de emisión y excitación de cada factor individualmente para los modelos con 3 y 4 factores, respectivamente. Junto a cada par de espectros se muestra el número total de cuentas por muestra, y la contribución relativa de ese factor al total de cuentas de cada muestra.

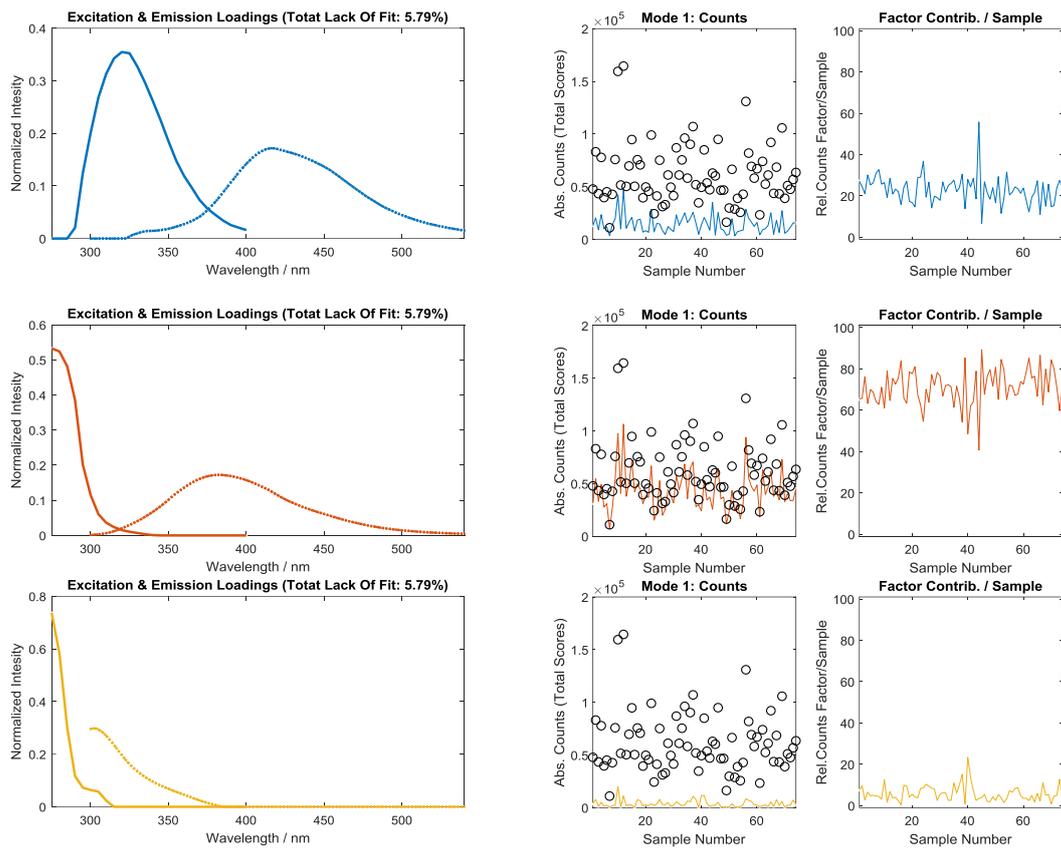


Figura 6.14 Espectros de excitación y emisión, cantidad de cuentas y contribución por muestra para cada uno de los factores del modelo PARAFAC con 3 factores.

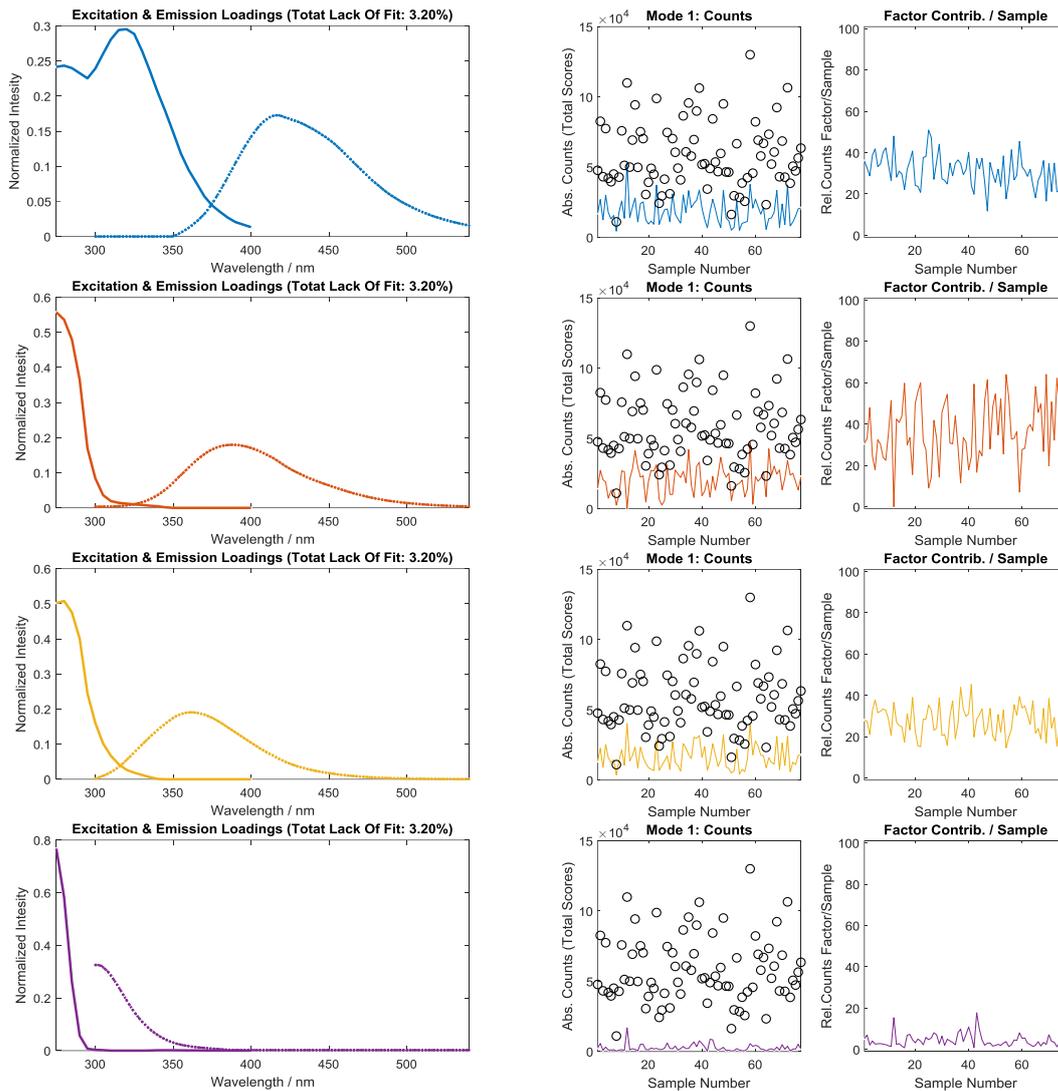


Figura 6.15 Espectros de excitación y emisión, cantidad de cuentas y contribución por muestra para cada uno de los factores del modelo PARAFAC con 4 factores.

Tal como puede observarse en las figuras, todos los espectros obtenidos cumplen las restricciones impuestas previamente (no negatividad para excitación y no negatividad y unimodalidad para los espectros de emisión). Además, en todos los casos, se verifica que el espectro de emisión está desplazado a mayores longitudes de onda que el de excitación correspondiente y que presentan una apropiada superposición entre ellos. De aquí que todos parecen ser físicamente aceptables. Sin embargo, la inspección del espectro de excitación del factor 1 correspondiente al modelo PARAFAC con 3 factores (figura 6.14) revela que el mismo toma valores de intensidad cero en las primeras longitudes de onda del rango, y luego sube abruptamente. Este comportamiento no es aceptable para un espectro de excitación. Además, en esa misma zona, los espectros de excitación de los factores 2 y 3 presentan valores altos de intensidad, que caen también abruptamente a partir de los 290 nm, que es justamente la zona donde el factor 1 incrementa su valor. Esta

correlación negativa entre el comportamiento del factor 1 y los otros dos factores en esa región espectral es un fuerte indicio de acoplamiento matemático durante el proceso de descomposición. Por este motivo, el modelo con 3 factores también fue descartado. Respecto de los espectros del modelo PARAFAC con 4 factores (figura 6.15), tanto los espectros de emisión como los de excitación obtenidos presentan formas apropiadas y no se observan irregularidades. En consecuencia, se consideró que los espectros obtenidos con el algoritmo son físicamente aceptables.

En síntesis, el modelo PARAFAC con 4 factores presenta espectros aceptables tanto en emisión como en excitación, y al mismo tiempo posee un porcentaje de falta de ajuste relativamente bajo y una consistencia del núcleo elevada. Esta es la situación de compromiso para este sistema, y de aquí se concluye que el número óptimo de factores para este sistema es 4.

6.2.2 Modelo PARAFAC obtenido

A partir del pre-procesamiento de las 100 MEE de muestras de orina descrito en la sección 6.2 y considerando que 4 factores es el número óptimo para este sistema, se obtuvo el modelo PARAFAC que se describe a continuación. Las longitudes de onda de los máximos de excitación y emisión correspondientes a cada uno de los factores resueltos por el algoritmo se listan en la tabla 6.2.

Factor	Longitud de onda del máximo en excitación	Longitud de onda del máximo en emisión
1	320 nm	420 nm
2	275 nm	388 nm
3	280 nm	362 nm
4	275 nm	300 nm

Tabla 6.2 Longitudes de onda de los máximos de emisión y excitación correspondientes a los 4 factores del modelo PARAFAC

Los máximos de los espectros de excitación de los factores 2, 3 y 4 se encuentran en longitudes de onda muy cercanas (entre 275 y 280 nm), por lo que podría pensarse que corresponden a un único factor. Sin embargo, sus respectivos espectros de emisión son bien diferentes entre sí y presentan sus máximos en distintas longitudes de onda. Esto puede observarse con mayor claridad en la figura 6.12, donde se presentaron los espectros de excitación de los cuatro factores superpuestos. Es importante señalar aquí que la zona espectral donde se agrupan los espectros de excitación de los factores mencionados, es justamente la misma en la que las muestras de orina presentan una banda de absorción

muy importante (figura 6.6). Por lo tanto, los perfiles de excitación obtenidos por el algoritmo son completamente razonables.

Las superficies de contorno correspondientes a cada uno de los factores obtenidos se muestran en la figura 6.16. Junto cada una de ellas, se representan los espectros de excitación y emisión correspondientes.

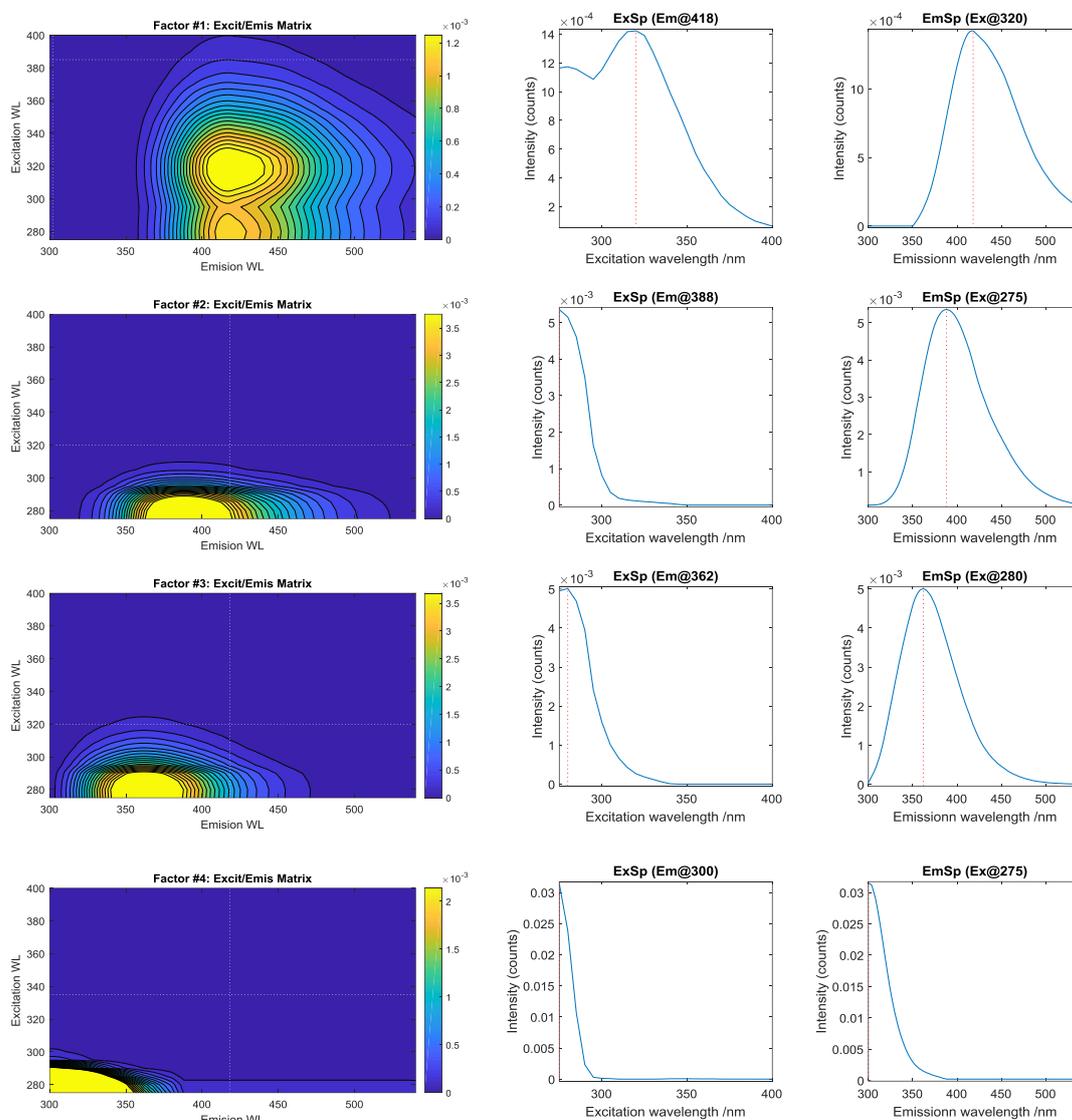


Figura 6.16 Superficie de contorno de los 4 factores del modelo PARAFAC (izq.), junto sus correspondientes espectros de excitación (centro) y emisión (der.).

Tal como se mencionó previamente, la falta de ajuste global del conjunto de todas las MEE estimadas por el modelo fue del 3,20%. Sin embargo, los porcentajes de falta de ajuste individuales para cada una de las MEE también fueron analizados. Las figuras 6.17 y 6.18 muestran gráficos representativos que dan las cotas mínima y máxima para los porcentajes de falta de ajuste obtenidos con el modelo PARAFAC de 4 factores.

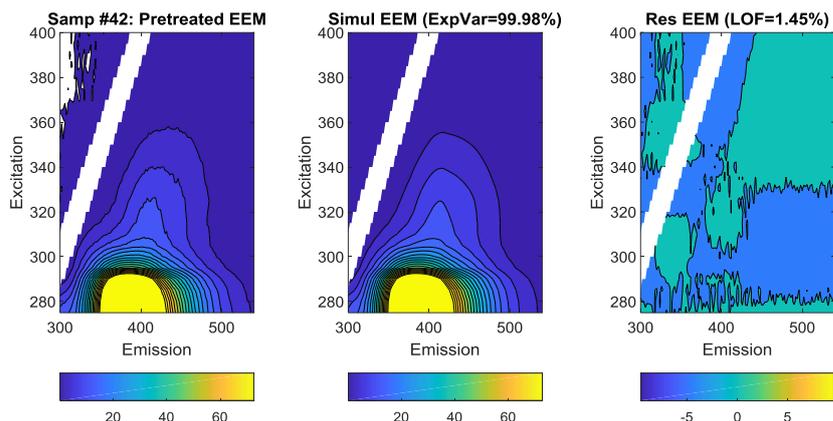


Figura 6.17 MEE medida experimentalmente (izq.), matriz calculada por PARAFAC (centro), y matriz de los residuos (der.) para la muestra que presentó el mejor ajuste (LOF 1,45%).

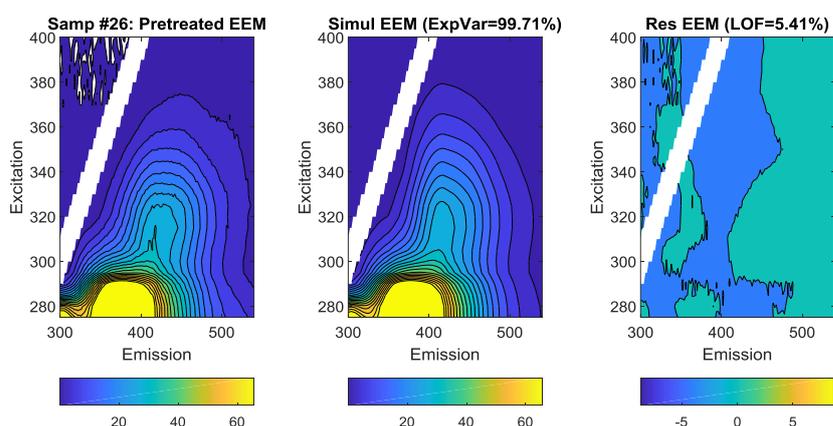


Figura 6.18 MEE medida experimentalmente (izq.), matriz calculada por PARAFAC (centro), y matriz de los residuos (der.) para la muestra que presentó el mejor ajuste (LOF 5,41%).

A pesar de que la figura 6.18 muestra una MEE cuyo porcentaje de falta de ajuste individual es mayor que el de la MEE de la figura la 6.17 (1,45% vs. 5,41%), en ambos casos los valores obtenidos son aceptables y además se verifica que los residuos no presentan ninguna estructura evidente.

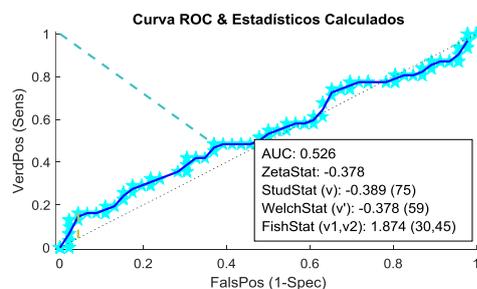
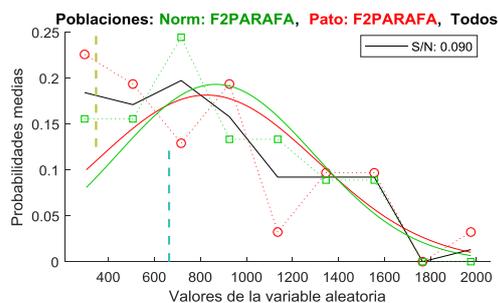
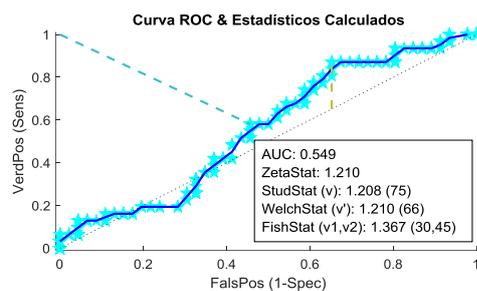
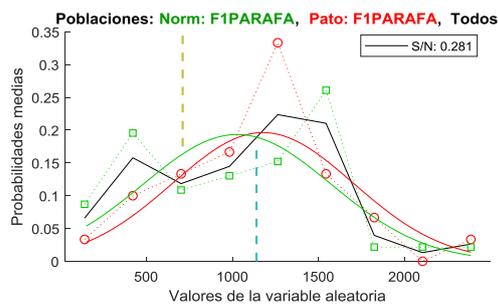
Los resultados de esta sección muestran que el algoritmo PARAFAC fue capaz de desarrollar un modelo trilineal de 4 factores que ajusta satisfactoriamente al arreglo de tres vías asociado al conjunto de MEE estudiadas. Por lo tanto las matrices A, B y C representan una adecuada síntesis de la información espectral registrada experimentalmente en el rango comprendido entre λ_{em} 300 a 540 nm y λ_{Ex} 275 a 400 nm.

6.3 Análisis estadístico de las MEE

De acuerdo con el objetivo de la segunda parte del presente trabajo de Tesis, se evaluaron las diferencias y similitudes entre las MEE correspondientes a muestras de orina provenientes de personas sanas y con enfermedades oncológicas. Teniendo en cuenta que el algoritmo PARAFAC fue capaz de descomponer satisfactoriamente el arreglo de tres vías conteniendo el conjunto de las MEE en tres matrices (A: scores, B: espectros de emisión y C: espectros de excitación), se estudió matriz de los scores (A) del modelo PARAFAC de la sección 6.2.2 mediante el software de análisis estadístico descrito en el Capítulo 5. De esta manera se consideraron 4 nuevos índices de interés, cada uno de los cuales corresponde a los scores de uno de los factores del modelo PARAFAC. Los valores estimados de los parámetros estadísticos para cada uno de estos índices se listan en la tabla 6.3, mientras que la figura 6.19 muestra los histogramas y curvas ROC correspondientes.

Índice	S/N	F-Fisher	t-Student	t-Welch	(f)	AUC	
I _{F1PARAFAC}	0,281	1,367 (30,45)	1,208 (75)	1,210	(66)	0,549	✗
I _{F2PARAFAC}	0,900	1,874 (30,45)	0,389 (75)	0,378	(59)	0,526	✗
I _{F3PARAFAC}	0,291	1,559 (30,45)	1,254 (75)	1,240	(63)	0,583	✗
I _{F4PARAFAC}	0,408	2,675 (45,30)	1,758 (75)	1,966	(70)	0,594	✗

Tabla 6.3 Valores de los estadísticos estimados para los índices derivados del modelo PARAFAC



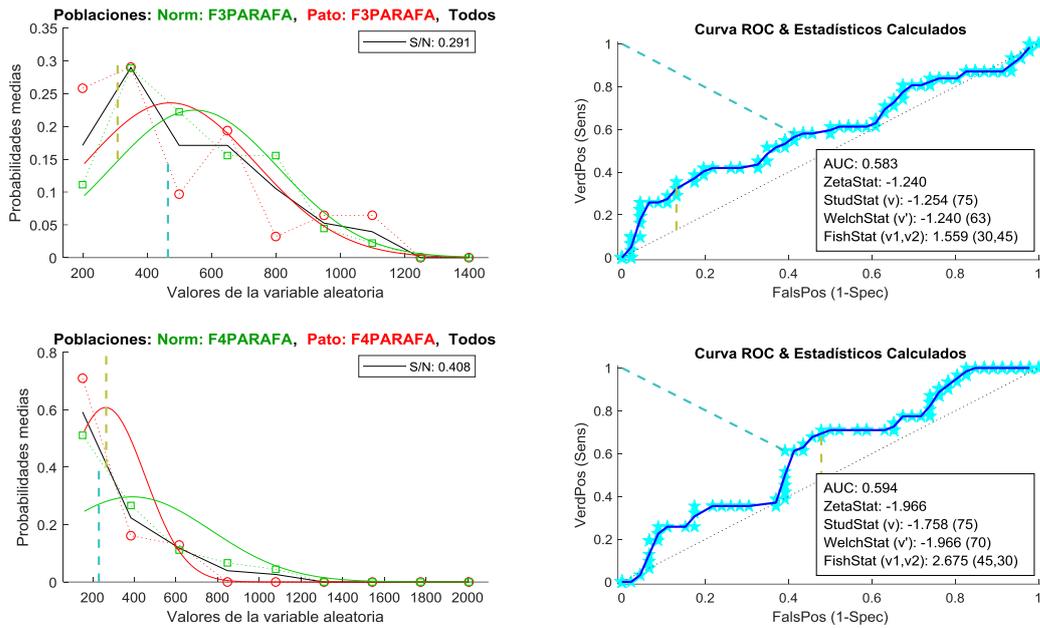


Figura 6.16 Histogramas (Izq.) y curvas ROC (Der.) correspondientes a los índices $I_{F1PARAFAC}$, $I_{F2PARAFAC}$, $I_{F3PARAFAC}$ e $I_{F4PARAFAC}$

Los valores estimados para la relación señal ruido en ningún caso superan la unidad, razón por la cual se considera que estos índices no contienen ningún tipo de información que pueda dar cuenta de la existencia de las dos poblaciones por sobre el nivel de aleatoriedad. Además, del análisis estadístico se desprende la misma conclusión: en ninguno de los 4 casos hay evidencia de que existan diferencias significativas entre las medias para sanos y enfermos: Los valores del estadístico F estimados para los índices $I_{F1PARAFAC}$, $I_{F2PARAFAC}$ e $I_{F3PARAFAC}$ presentaron valores menores al $F_{crítico}$ ($F^{0,01}_{30,45}=2,14$), por lo que se asumió la homogeneidad de las varianzas y se evaluó la diferencia entre las medias poblacionales con el test de Student, mientras que el valor de F estimado para el $I_{F4PARAFAC}$ fue mayor a su $F_{crítico}$ correspondiente ($F^{0,01}_{45,30}=2,23$) por lo que la diferencia entre las medias poblacionales se estudió con el test de Welch. Para los cuatro índices, el valor del estadístico de t obtenido fue menor al t crítico correspondiente según los grados de libertad correspondientes a cada caso.

En este punto es importante señalar que si bien el modelo PARAFAC permitió un excelente ajuste de las MEE, los índices obtenidos no resultaron apropiados para discriminar entre ambas poblaciones. Este resultado parece indicar que la metodología tal como fue aplicada en el presente Capítulo no es apropiada a los fines de desarrollar un test diagnóstico. Por lo tanto, del análisis expuesto se concluye que en la zona espectral analizada la información contenida, aunque bien resuelta por el algoritmo PARAFAC, no es suficiente como para poder discriminar las orinas pertenecientes a personas que padecen enfermedades oncológicas de las que no.

6.3.1 Estudio de la zona espectral óptima de las MEE para desarrollar un test diagnóstico

En la sección anterior se mostró que el modelo PARAFAC ajustó muy bien las MEE en la zona espectral en la que se encuentran la mayoría de los fluoróforos de interés biológico. Sin embargo, el procedimiento presentado en este Capítulo no logró discriminar las muestras de orina que correspondían a personas sanas de aquellas asociadas a personas afectadas con alguna patología oncológica. Por otro lado en el Capítulo 5, a partir de los espectros emisión de fluorescencia registrados mediante excitación con luz de 405 nm, se obtuvieron índices sumamente eficientes en dicha discriminación. Entonces, para estudiar la relación entre ambos enfoques se realizó un análisis de la información contenida en las MEE del presente Capítulo en la zona espectral que se corresponde con la de los espectros estudiados en el Capítulo 5. Es decir, se extrajeron los espectros de emisión de fluorescencia a partir de las MEE. Para este análisis se empleó un subconjunto de MEE medido en un rango extendido, que abarcaba longitudes de onda de excitación de hasta 450 nm. Dicho subconjunto estaba formado por 63 MEE, 36 de las cuales correspondían a muestras de orina de personas sanas y 27 a personas con patologías oncológicas. Se extrajeron entonces las “rodajas” de las MEE correspondientes a excitaciones de 395, 400, 405, 410 y 415 nm. De esta manera se generaron 5 conjuntos de datos de 63 muestras cada uno. De ellos, el que corresponde a la excitación con 405 nm se muestra en la figura 6.20.

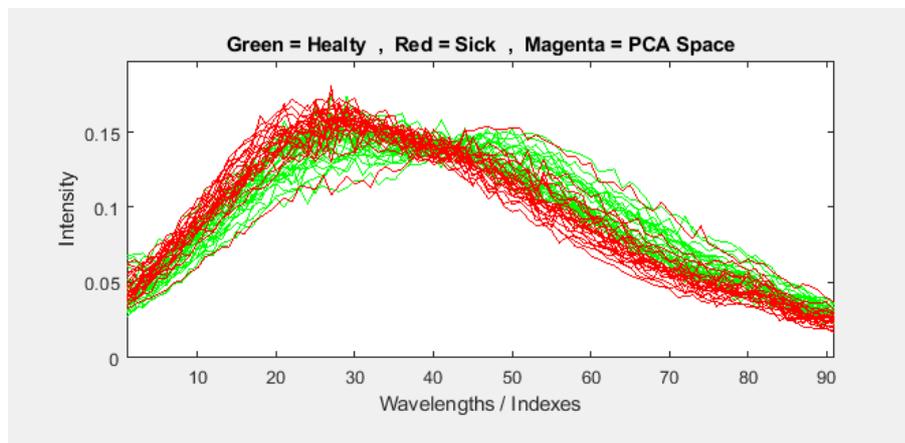


Figura 6.20 Espectros correspondientes a la excitación con 405 nm extraídos de las MEE

Tal como puede observarse, en general los espectros correspondientes a las personas con patologías oncológicas (en color rojo) se encuentran ligeramente desplazados a la izquierda respecto de los correspondientes a muestras de personas sanas (en color verde). Esto es consistente con lo observado a partir de los espectros presentados en el Capítulo 5. Sin embargo, si bien aquí se llega a apreciar la existencia de las dos poblaciones, las señales de fluorescencia registradas son relativamente bajas y en consecuencia la relación S/N no es buena. Esto es consecuencia de los altos factores de

dilución empleados en el pre tratamiento de las muestras que se estudiaron en este Capítulo. Cabe recordar que dicha dilución fue necesaria para poder registrar apropiadamente la fluorescencia en la región UV, teniendo que diluir entre 150 y 700 veces las muestras para evitar pérdidas de linealidad en las señales. Sin embargo, resulta evidente que en el rango espectral visible esta dilución repercute negativamente en la capacidad para discriminar entre pacientes sanos y enfermos ya que disminuye sustancialmente la relación señal/ruido.

Asimismo, con fines comparativos, a cada conjunto de datos extraído de las MEE correspondiente a una longitud de onda de excitación se analizó por PCA de manera similar al Capítulo 5. En la tabla 6.4 se resumen los resultados obtenidos de dicho análisis y en la figura 6.21 se muestra una imagen representativa del análisis por PCA.

λ_{Ex}	Outliers	FN	FP	Precisión
395 nm	7	5	11	0,7143
400 nm	7	5	10	0,7321
405 nm	7	5	10	0,7321
410 nm	7	6	13	0,6607
415 nm	5	5	12	0,7069

Tabla 6.4 Resultados del análisis por PCA de los espectros extraídos de las MEE

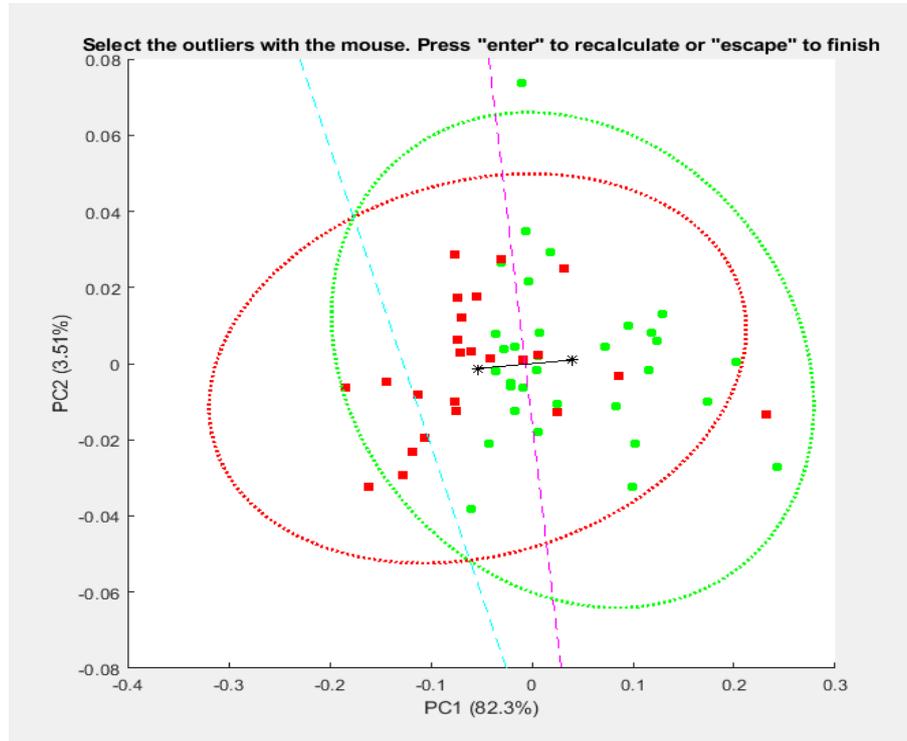


Figura 6.21 Puntos en el espacio PC1/PC2, con elipses de confianza ($1-\alpha=99\%$), calculadas a partir de los EEm ($\lambda_{ex}=405$ nm) de las muestras de orina de personas sanas (en verde) y de las diagnosticadas con alguna patología oncológica (en rojo).

Del análisis expuesto se concluye que si bien la información correspondiente al rango λ_{ex} 395-415 nm contenida en las MEE medidas es de baja calidad, presenta evidencia de la existencia de dos poblaciones. La baja calidad de la información se atribuye entonces a la gran dilución realizada en el pre-procesamiento de las muestras, y se manifiesta en los espectros con valores de intensidad de fluorescencia muy bajos.

La región espectral UV es donde se registra la mayor intensidad de fluorescencia de la orina, ya que la gran mayoría de los componentes biológicos presentes en la misma fluorescen en ese rango (ver figura 6.22).

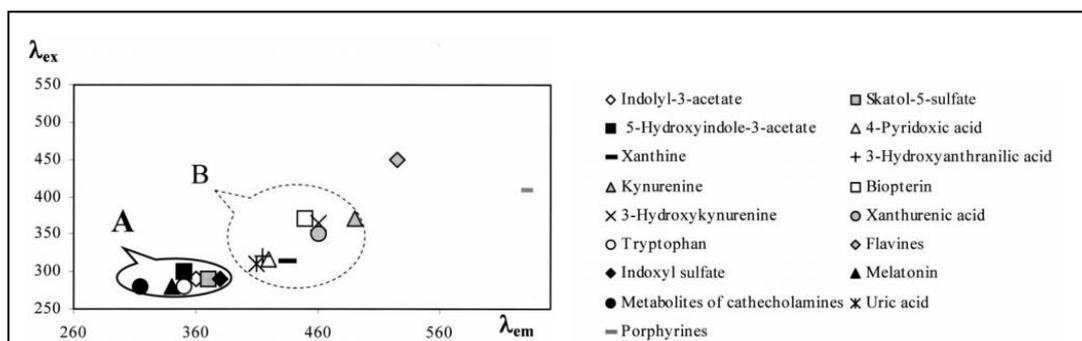


Figura 6.22 Localización de los fluoróforos presentes en la orina. El grupo A de fluoróforos representa la mayor contribución a la fluorescencia total de la orina. El grupo B es característico con 4 a 6 veces menor concentración que el A. Adaptado de [82].

A medida que el rango espectral considerado se corre hacia mayores longitudes de onda, la cantidad de analitos capaces de presentar fluorescencia es mucho menor (ver figuras 6.17, 6.18 y 6.22). En consecuencia, para poder registrar correctamente la fluorescencia en la zona visible se requiere una mayor concentración de la muestra (o menor dilución) que en el rango UV. Se puede interpretar entonces que las MEE analizadas en la sección 6.3 contienen gran cantidad de información, pero en una zona espectral en la que esa información no es útil a los fines de discriminar entre las poblaciones de pacientes sanos y enfermos.

Del análisis realizado se desprende que la zona espectral en la que se ven alteradas las señales de fluorescencia de la orina de pacientes afectados por patologías oncológicas es aquella que corresponde a los metabolitos de la región superior derecha del grupo B en la figura 6.22. Por otra parte, también se puede afirmar que la región espectral asociada a la fluorescencia de los metabolitos del grupo A no se ve significativamente comprometida en personas que cursan procesos oncológicos respecto de las que no.

Por lo tanto, a partir de los resultados obtenidos en los Capítulos 5 y 6, resulta evidente que la información de mejor calidad a los fines de desarrollar un test diagnóstico se encuentra en la zona visible, y no en la región UV del espectro.

Capítulo 7. RESUMEN, DISCUSIÓN Y CONCLUSIONES GENERALES

El presente trabajo de Tesis se ha enfocado en la realización de investigaciones tendientes a evaluar la factibilidad del empleo de métodos ópticos no invasivos como herramientas para el diagnóstico y monitoreo del tratamiento de enfermedades oncológicas

Discusión de los resultados de la Parte I

El objetivo de la primera parte fue evaluar el potencial de las observaciones ópticas “*in vivo*” para el diagnóstico de neoplasias de tejido humano empleando técnicas de fluorescencia sobre la zona tumoral y sus adyacencias. A continuación se presenta el análisis conjunto de los espectros de emisión de fluorescencia (EEm) registrados en los casos de estudio abordados en el Capítulo 3. En todas las figuras se muestran los espectros correspondientes a zonas sanas a la izquierda, mientras que del lado derecho se muestran los correspondientes a zonas afectadas.

La figura 7.1 muestra los espectros de todos los tejidos estudiados en el Capítulo 3, escalados a norma euclidiana unitaria. Los espectros de cada uno de los casos se presentan con un color de línea diferente y dicho color se mantiene para los gráficos de zonas sanas y zonas afectadas.

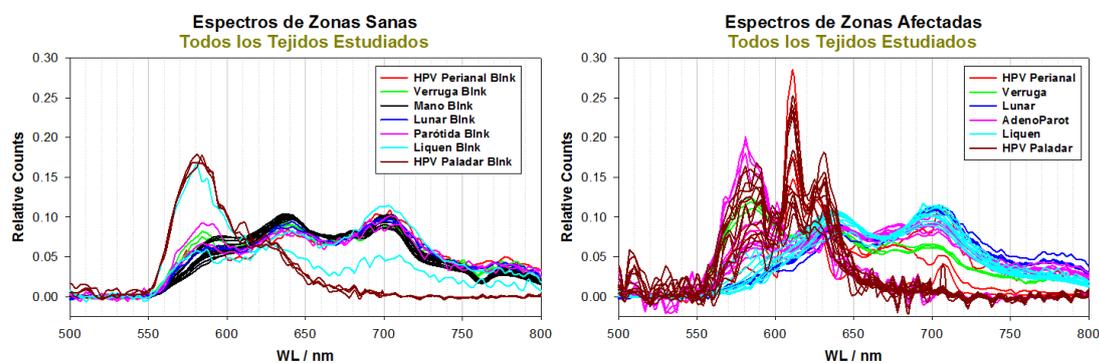


Figura 7.1 Espectros de todos los tejidos estudiados

El análisis de la figura 7.1 revela que los espectros correspondientes a zonas sanas estudiadas parecen estar englobados todos dentro de dos estructuras bien definidas, mientras que los de las zonas afectadas presentan una gran variabilidad. Sin embargo, dado que los espectros corresponden a patologías diferentes y a tejidos diferentes, con el fin de realizar un análisis un poco más detallado se graficaron los espectros agrupados según pertenezcan a casos similares.

En la figura 7.2 se muestran sólo los espectros que fueron tomados sobre la piel. Es decir, de todos los casos estudiados en el Capítulo 3, se excluyeron aquellos que fueron registrados dentro de la cavidad bucal. De esta manera se obtiene la figura de la izquierda para las zonas de piel sana, y la de la derecha que corresponde a las zonas de piel afectadas por alguna patología.

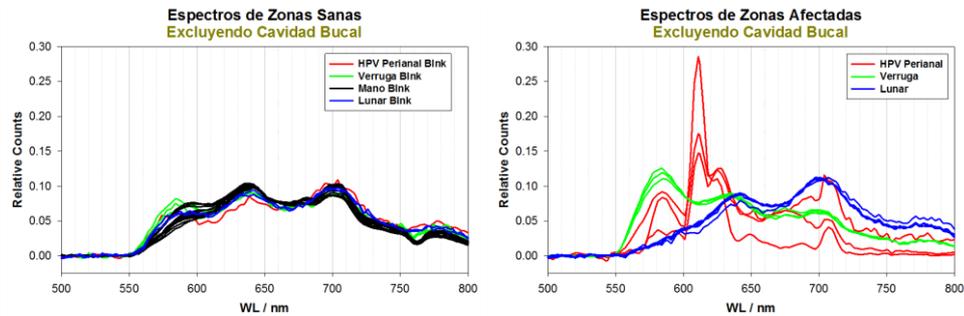


Figura 7.2 espectros de todos los tejidos estudiados sobre piel

Puede observarse que los todos espectros registrados en zonas sanas de la piel, escalados a norma euclídea son prácticamente iguales, independientemente de la zona del cuerpo a la que corresponde dicho epitelio. Es decir, los espectros de fluorescencia natural registrados sobre la piel de distintas personas, en zonas bien diferentes del cuerpo como son la mano, el surco nasogeniano, la articulación acromioclavicular, el hombro y la región perianal, tienen la misma estructura. Se observa, además, que todos ellos no sólo presentan las mismas bandas, sino que también conservan la relación entre sus intensidades. En contraste a éstos, los espectros de las zonas afectadas muestran estructuras bien diferentes dependiendo del caso, tal como se observa en el panel derecho de la figura 7.2.

Por último, resultó de interés agrupar los espectros según la patología que los afectaba, en lugar de hacerlo según el tipo de tejido. Entonces, se graficaron juntos los espectros de los dos casos de pacientes con HPV6, descritos previamente. Dichos espectros se muestran en la figura 7.3.

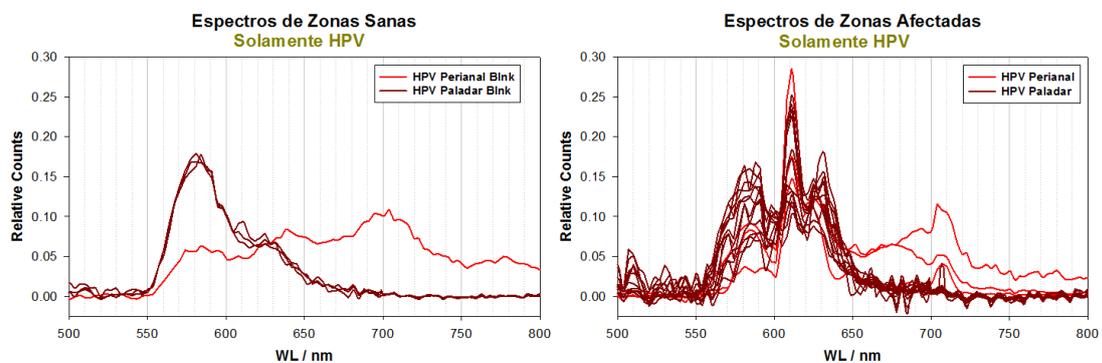


Figura 7.3 espectros de todos los tejidos estudiados

En el panel de la derecha se observa que los espectros de las zonas afectadas por el HPV6 presentan un pico muy pronunciado en 611 nm, que los diferencia del resto de los espectros. Asimismo, en ambas patologías, a los lados de ese pico se observan bandas de menor intensidad que tienen características bien diferentes a las de las bandas asociadas a otras patologías, e incluso a las de tejidos sanos. Los espectros que corresponden a zonas sanas se muestran en el recuadro izquierdo de la misma figura. Tal como puede observarse, cada zona presenta su espectro característico, según el tipo de tejido al que pertenece. Es muy importante resaltar que, a pesar de que los espectros correspondientes a los tejidos sanos son bien diferentes para la piel y la mucosa bucal, ambos tejidos cuando se encuentran afectados por la infección por HPV presentan espectros similares con patrones que se repiten en ambos casos.

A partir de estos ejemplos, en principio, se puede pensar en realizar una generalización de lo observado. Sin embargo, luego de haber analizado todo el material obtenido a partir de las biopsias ópticas realizadas por fluorescencia sobre los casos clínicos investigados y de haber analizado en forma crítica tanto la cantidad como la calidad de la información que dicho material ha brindado, consideramos que es necesario aumentar la cantidad de casos a estudiar antes de establecer conclusiones categóricas. Sin embargo, estos resultados preliminares son muy importantes y constituyen una prueba de la potencia de la espectroscopia óptica de fluorescencia como técnica de diagnóstico.

Discusión de los resultados de la parte II

El objetivo de la segunda parte del trabajo de Tesis fue contribuir al desarrollo de un método diagnóstico que, a partir del análisis de la autofluorescencia de la orina, pueda ser utilizado como método de “screening” para la detección de enfermedades neoplásicas. En este contexto, como parte del estudio se han investigado cientos de índices generados a partir de los espectros de fluorescencia de las muestras de orina con el fin de establecer aquellos que mejor rendimiento tienen en cuanto al poder de discriminación entre personas que padecen enfermedades oncológicas y personas que no. En el Capítulo 5 se presentó una gran cantidad de esos índices, comenzando por los más sencillos (como intensidades de fluorescencia registradas a una determinada longitud de onda, o cocientes entre ellas), y finalizando con aquellos que implican una complejidad matemática superior o un mayor nivel de pre-procesamiento. En muchos casos se encontró, sin embargo, que los resultados obtenidos a partir de estos índices no estuvieron directamente relacionados a su complejidad.

En la sección 5.3 se estudiaron los índices generados a partir de un valor único asociado cada espectro, como por ejemplo la intensidad de fluorescencia a una determinada longitud de onda o el simple cociente o la diferencia entre dos de ellas. De

todos ellos, los listados en las tablas 5.2 y 5.3 son los que mostraron tener la capacidad de discriminar las poblaciones de sanos y enfermos, y el que presentó el mayor poder de discriminación fue el índice $I_{471/507}$. Luego, en la sección 5.7 se aplicó el análisis de discriminante lineal (LDA) al conjunto de todos los índices que mostraron capacidad discriminatoria en la sección 5.3. Sin embargo, el rendimiento de este nuevo test no presentó grandes mejoras respecto del índice $I_{471/507}$ individual.

A partir de la sección 5.4 se estudiaron índices generados mediante procedimientos matemáticos más complejos, en el sentido de que pudieran abarcar mayor volumen de información del espectro original. Entonces, en un primer análisis se realizó la reducción de la “dimensionalidad” de los espectros completos mediante PCA, y se analizaron los scores de los ocho primeros componentes como índices de interés. El resultado fue contundente: sólo el primer componente principal (PC1) presentó una diferencia significativa entre los valores medios de las poblaciones de pacientes sanos y enfermos. Los PCs del número 2 en adelante mostraron no ser útiles a tal fin. Sin embargo, luego de analizar con el algoritmo LDA al conjunto formado por los primeros ocho PCs, el rendimiento del test (cuantificado a través del AUC), fue un 3,7% superior al obtenido empleando solo PC1 como índice. Por lo tanto, en este caso, la incorporación de los siete PC descartados originalmente y la implementación del LDA permitió aumentar significativamente la eficiencia en la clasificación. En otras palabras, la implementación del PCA como herramienta de pre-procesamiento para reducir el volumen de datos de la matriz original y el posterior análisis por LDA resultó en el desarrollo de un test más eficiente en la discriminación de poblaciones, siendo incluso un 1,36% superior que el mejor de los índices estudiados en la sección 5.3.

En la sección 5.5 se estudió si la implementación de la función derivada como pre-procesamiento de la información espectral servía para generar algún índice cuyo rendimiento superara a los precedentes. Se propuso esta operación dada la capacidad de la función derivada para resaltar pequeñas variaciones de los espectros. Los resultados obtenidos indicaron que la función derivada de orden uno mejora un 0,5% el rendimiento del test para índices individuales respecto del mejor de los índices generados a partir de los espectros sin derivar. De aquí que de los índices asociados a una sola longitud de onda o a un pequeño intervalo del espectro, la primera derivada fue la que presentó los índices de mayor discriminación. Luego, el análisis por LDA (sección 5.7) realizado al conjunto de los 8 índices seleccionados de los espectros sin derivar y de sus derivadas de primer y segundo orden no presentó mejoras en cuanto al rendimiento global del test. Es decir, si bien la potencia en clasificación obtenida mediante el análisis por LDA fue comparable a las mayores obtenidas en el análisis realizado en la sección 5.5, no se lograron mejoras sustanciales respecto de los índices con mayor poder de clasificación. Por último, de la comparación de los valores obtenidos del análisis por LDA a los espectros derivados con los resultados de las secciones anteriores, se concluye que al incorporar las derivadas de

los espectros al análisis por LDA no se obtuvieron mejoras en la capacidad discriminatoria ya que ninguno de los test que incorporan la función derivada presentó mejor rendimiento respecto del obtenido a partir del pre-procesamiento con PCA y posterior aplicación del algoritmo LDA.

En la sección 5.6 se analizaron los scores de cada una de las cuatro componentes con las que se ajustó la matriz de los 81 espectros originales mediante MCR-ALS. El análisis de los 4 componentes individuales reveló que los componentes 1 y 2 presentan diferencias significativas entre los valores medios de las poblaciones de sanos y enfermos, mientras que los componentes 3 y 4 no mostraron ser útiles a los fines de detectar la presencia de dos poblaciones. De los dos primeros, el que mejor rendimiento presentó fue el componente 2, que fue un 9,17% superior al componente 1. También se analizó el índice formado por la diferencia entre los scores de los componentes 1 y 2. Este nuevo índice presentó un rendimiento apenas superior al del MCR2 individual. Finalmente, se analizaron los scores de los cuatro componentes juntos, mediante LDA. Los resultados obtenidos indicaron un aumento en el poder de clasificación del 2,12% respecto del mejor de los cuatro índices analizados individualmente en la sección 5.6. Sin embargo, el rendimiento global no superó al obtenido mediante la secuencia PCA+LDA.

Por último, y siguiendo con el objetivo de encontrar las características o índices de los espectros de fluorescencia de las muestras de orinas que mejor clasifiquen cada muestra según pertenezca a una persona sana o afectada por una patología oncológica, se usó LDA con nuevo conjunto de datos, formado por la totalidad de los índices analizados en detalle en el Capítulo 5. Así, en la sección 5.7 se estudió una matriz de datos de 81 muestras y 28 variables, donde cada variable era uno de los índices estudiados previamente. Las proyecciones sobre el eje de Fisher resultantes se muestran en la Figura 7.4. Puede observarse que este nuevo índice (denominado I_{LDA-28}) tiene la capacidad de discriminar muy bien entre individuos pertenecientes cada una de las poblaciones.

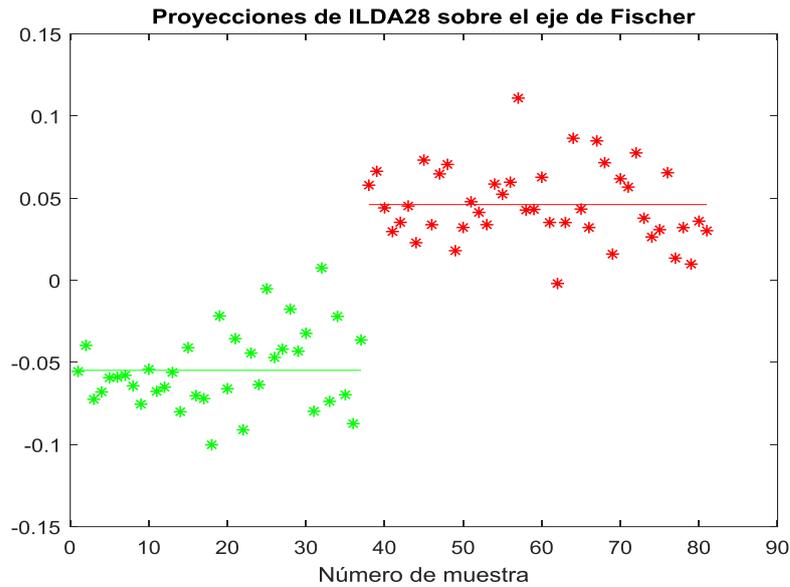


Figura 7.4 Proyecciones de I_{LDA-28} sobre el eje de Fisher

Cabe destacar que a partir de la aplicación del análisis estadístico se obtuvo el mismo resultado, por lo tanto se puede afirmar que el índice formado por los valores de las proyecciones sobre el eje de Fisher para el conjunto completo de datos de la tabla 5.7 (I_{LDA-28}) tiene la capacidad de clasificar a las muestras de orina, según sean de personas sanas o enfermas, ya que sus valores corresponden a distribuciones de probabilidad con medias poblacionales distintas.

Con el fin de evaluar la posibilidad de desarrollar un test dicotómico basado en el índice I_{LDA-28} para discriminar entre individuos de cada población se realizó la curva ROC correspondiente. La curva ROC obtenida (Fig 7.5) presentó una gran similitud con la curva ROC "ideal", o de clasificación perfecta, ya que roza la coordenada (0,1) del espacio ROC.

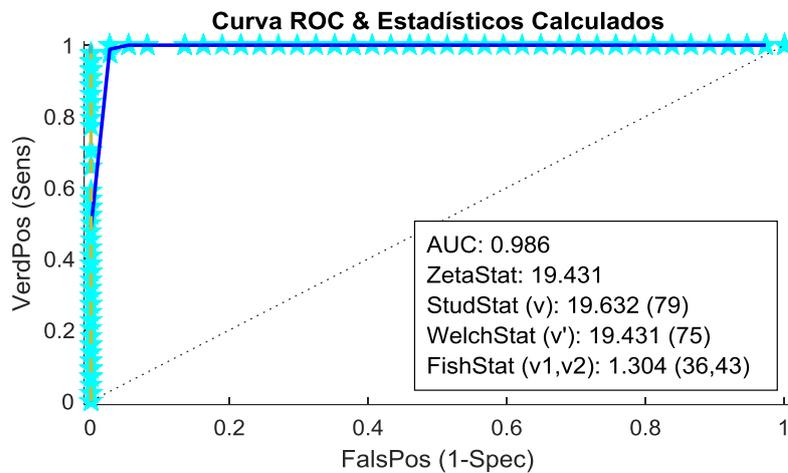


Figura 7.5 Curva ROC correspondientes las proyecciones en el eje de Fisher de los índices I_{LDA-28}

El AUC obtenido fue 0,986, que indica un excelente rendimiento diagnóstico. Esto quiere decir que existe un 98,6% de probabilidad de que ante un par de individuos, uno enfermo y el otro sano, la prueba los clasifique correctamente. Además, de la comparación con los AUC calculados para los otros test en el mismo Capítulo, este nuevo índice presenta un valor 1,86% superior al mejor obtenido (mediante el pre-procesamiento con PCA y posterior LDA).

A partir del estudio presentado en el Capítulo 5, se concluye que de todos los índices testeados el que mostró la mejor capacidad para discriminar ambas poblaciones es el obtenido por LDA a partir de los valores de la tabla 5.7 (I_{LDA-28}).

Otro resultado muy importante obtenido a partir del estudio de los espectros de emisión de orinas presentado en el Capítulo 5 se describió en la sección 5.4. En esta parte del trabajo se analizó la variabilidad manifiesta en los espectros de fluorescencia de orina asociada a las condiciones de toma de la muestra. Esta variabilidad “intra-individuo” resultó ser mucho menor a la variabilidad total “inter-individuo” tanto para la población total como para la de personas sanas. Resumiendo, de este análisis surgieron dos conclusiones muy importantes:

- 1) A pesar de que la composición de la orina varía permanentemente, la variabilidad de su fluorescencia intra-individuo es mucho menor a la variabilidad de la población total (inter-individuo), por lo que una metodología diagnóstica basada en el análisis de la fluorescencia de muestras de orina puede ser viable.
- 2) Para minimizar la variabilidad intra-individuo, y mejorar el rendimiento del test, es siempre conveniente analizar la primer orina de la mañana, en ayunas y con suficiente tiempo de retención.

El punto 1) es crucial y da sentido a las conclusiones que surgen de la evaluación de los pacientes considerados en el presente estudio.

Las investigaciones presentadas en el Capítulo 5 demuestran que parte de la información contenida en los espectros de fluorescencia de muestras de orina se puede utilizar para clasificar personas que padecen enfermedades oncológicas, respecto de las que no. Sin embargo, es importante considerar que no cualquier característica de los espectros posibilita un diagnóstico exitoso. Resulta evidente, entonces, la necesidad de aplicar herramientas matemáticas robustas para el análisis de la información espectroscópica, y a partir de allí considerar a la espectroscopia de fluorescencia como una técnica de “screening” adecuada para el diagnóstico.

Dados los excelentes resultados obtenidos en el Capítulo 5, y sabiendo que la mayoría de los analitos fluorescentes contenidos en la orina presenta una intensa emisión

en la región espectral UV, en el Capítulo 6 se amplió el rango espectral de análisis y además se estudiaron las denominadas matrices de emisión/excitación (MEE). En consecuencia se registraron las MEE, empleando longitudes de onda de excitación de 250 a 500 nm y de emisión entre 275 y 700 nm, de manera de evaluar la emisión de la mayor cantidad de los fluoróforos presentes en la orina y profundizar el análisis realizado sobre los espectros. Sin embargo, para las MEE medidas se observó que en la región de longitudes de onda por debajo de 380 nm la emisión era muy pobre y prácticamente no se registraba señal. Del análisis de los espectros de absorción correspondientes se concluyó que este efecto era consecuencia directa de los efectos de filtro interno (IFE) asociados a las propias muestras de orina, por lo que surgió la necesidad de pre-tratar las muestras diluyéndolas. En este contexto, se incrementó el factor de dilución a valores de entre 150 y 700 (dependiendo de la muestra en particular), para evitar pérdidas de linealidad en las señales medidas. En estas condiciones se registraron 100 MEE de muestras de orina, 48 de las cuales correspondían a personas sanas y 52 a personas con patologías oncológicas.

En línea con el objetivo planteado en la segunda parte del presente trabajo de Tesis, se buscó la manera de evaluar las diferencias y similitudes entre las MEE correspondientes a muestras de orina provenientes de personas sanas y con enfermedades oncológicas. Cabe destacar que la interpretación de la información contenida en las MEE del Capítulo 6 reviste mayor complejidad que en el caso de los espectros del Capítulo 5 debido al incremento de la dimensión de los datos. Con el fin de sintetizar la información espectral registrada experimentalmente, se evaluó la capacidad del algoritmo PARAFAC para generar un modelo que reproduzca la información contenida en las MEE medidas experimentalmente. La construcción del modelo PARAFAC para este conjunto de datos se presentó en la sección 6.2. Allí se describió en detalle el pre-procesamiento de las señales registradas, la exploración del conjunto de datos y el desarrollo del modelo. Luego de un detallado análisis, los resultados de la sección 6.2.2 revelan que algoritmo PARAFAC fue capaz de construir un modelo de 4 factores que ajusta satisfactoriamente al conjunto de MEE medidas experimentalmente en el rango comprendido entre λ_{Em} 300 a 540 nm y λ_{Ex} 275 a 400 nm.

Posteriormente, se evaluó la capacidad de las MEE para discriminar entre las muestras de orina de personas sanas y de pacientes afectados a través de su representación según el modelo PARAFAC. Dado que el modelo de 4 factores representó adecuadamente al conjunto de las MEE, se usaron los scores del modelo (matriz **A**) para analizar el significado estadístico de la información contenida en las MEE. De esta manera se generaron 4 nuevos índices de interés, que corresponden a los scores de cada uno de los factores del modelo PARAFAC para luego, en la sección 6.3, estudiarlos mediante el software de análisis estadístico descripto en el Capítulo 5. El análisis reveló que ninguno de los 4 índices presenta evidencia acerca de la existencia de diferencias significativas entre las medias para pacientes sanos y enfermos y, por lo tanto, ninguno de ellos podría ser útil

a los fines de desarrollar un test diagnóstico. En resumen, si bien el modelo PARAFAC permitió un ajuste muy satisfactorio de las MEE, los índices obtenidos no resultaron apropiados para discriminar entre ambas poblaciones.

Con el fin de estudiar la relación entre los resultados obtenidos a partir de los EEm estudiados en el Capítulo 5 y de las MEE investigadas en el Capítulo 6, se realizó un análisis de la zona espectral de las MEE que se corresponde con la de los EEm. Es decir, se extrajeron los espectros de emisión de fluorescencia a partir de las MEE. Los estudios realizados sobre estos “espectros” se presentaron al final del Capítulo 6 y mostraron que, a causa del incremento de dos órdenes de magnitud en el factor de dilución, la señal fluorescente registrada a partir de la excitación con longitudes de onda cercanas a 405 nm resultó de muy baja calidad. Sin embargo, a pesar de ello, presentó evidencia de la existencia de dos poblaciones.

Este resultado parece indicar que la metodología de registro y procesamiento de las EEM tal como fue aplicada en el Capítulo 6 no resulta adecuada a los fines de desarrollar un test diagnóstico. Por lo tanto, del análisis expuesto se concluyó que en la zona espectral analizada la información contenida, aunque bien resuelta por el algoritmo PARAFAC, no es suficiente como para poder discriminar las orinas pertenecientes a personas que padecen enfermedades oncológicas de las que no.

A partir de los resultados obtenidos en los Capítulos 5 y 6, resulta evidente que la información de mejor calidad a los fines de desarrollar un test diagnóstico está en la zona visible, y no en la región UV del espectro.

Conclusiones

La primera conclusión que surge del trabajo realizado es que la fluorescencia de la orina de los pacientes con cáncer exhibe diferencias espectrales en comparación con la observada para pacientes normales, y que esas diferencias pueden potenciarse mediante la aplicación de diferentes técnicas numéricas para el análisis de datos con el objeto de generar herramientas para el diagnóstico de procesos oncológicos. En ese sentido, en el presente trabajo de Tesis se han realizado importantes avances en cuanto al desarrollo y optimización de un protocolo de trabajo para la formulación de un test no invasivo, rápido y económico que permita la detección temprana de procesos oncológicos. Cabe señalar que, desde el punto de vista experimental, se analizaron exhaustivamente las condiciones de toma muestras, su conservación y preparación para establecer estrategias apropiadas que garanticen el mejor resultado del test. Por otro lado, es importante destacar que gracias a la aplicación conjunta de técnicas estadísticas, herramientas de cálculo numérico y métodos del álgebra matricial para el análisis de la información espectroscópica, se han logrado importantes avances respecto del desarrollo de parámetros con capacidad

diagnóstica (índices). En este contexto, se evaluaron múltiples transformaciones matemáticas de los datos espectroscópicos y se generaron más de trescientos índices de interés, que fueron analizados uno a uno, y posteriormente clasificados según su capacidad diagnóstica (capacidad de discriminar entre individuos pertenecientes a las poblaciones de sanos y enfermos).

A partir de las investigaciones desarrolladas surge como conclusión más importante que la aplicación de la espectroscopia de fluorescencia de muestras de orina puede ser considerada como una herramienta de diagnóstico para la detección de procesos oncológicos. Los resultados obtenidos han mostrado una exactitud global de al menos un 98%. Además, esta técnica posee una serie de ventajas por sobre las metodologías de diagnóstico existentes. En particular, resulta económica, muy sensible y fundamentalmente no invasiva. En consecuencia la consideramos sumamente apropiada para ser aplicada como herramienta de detección masiva (screening).

Es interesante señalar que el estudio de la variabilidad asociada a la fluorescencia de la orina de un mismo paciente, colectada en diferentes condiciones, muestra claramente que la aplicación de las herramientas de análisis desarrolladas en el presente Trabajo de tesis podría arrojar resultados aún mejores si se consideraran muestras de orina que cumplan condiciones más estrictas al momento de su colección.

Asimismo, cabe destacar que como parte del presente trabajo de Tesis se han generado valiosas herramientas de software que han permitido agilizar sustancialmente el análisis la información espectroscópica. El conjunto de rutinas desarrolladas incluye funciones para: el análisis estadístico de los diferentes índices, la reducción de dimensionalidad mediante el empleo de componentes principales (PCA) y la generación de curvas ROC para comparar la potencia de diferentes tests dicotómicos.

Finalmente, debe tenerse en cuenta que se ha trabajado con una cantidad de muestras suficiente como para obtener resultados con valor estadístico. A partir de este conjunto de muestras se han logrado importantes avances que permitieron el desarrollo de un protocolo preliminar para poder formular una sistemática de “screening” de eventuales pacientes patológicos.

Índice de abreviaturas utilizadas:

5-ALA: ácido d-aminolevulínico

AUC: (de Area Under Curve) área bajo la curva

AV: (de average scan) cantidad de escaneos promedio

BG: (de background) blanco absorbente que no emite señal del dispositivo de calibración del espectrofluorómetro

DFD: diagnóstico fotodinámico

EEm: espectro de emisión de fluorescencia

EEx: espectro de excitación de fluorescencia.

FG: (de fluor glass) vidrio fluorescente del dispositivo de calibración del espectrofluorómetro portátil

FS: fotosensibilizador

IFE: corrección por filtro interno

IT: tiempo de integración

LDA: Análisis discriminante lineal

LOF: (de Lack of fit), Falta de ajuste

MCR: resolución multivariada de curvas

MEE: matrices de emisión y excitación

PCA: análisis de componentes principales

PC_n: "n"-ésima componente principal

PpIX: protoporfirina IX

ROC: iniciales del término inglés original: Receiver Operating Characteristics

S/N: (de signal to noise), relación señal/ruido

Sup./Inf.: superior / inferior, respectivamente

TFD: terapia fotodinámica

BIBLIOGRAFIA

- [1] <http://www.diccionariomedico.net/>
- [2] Ferrer-Roca, Olga. "Endomicroscopia en anatomia patologica. Biopsia óptica." *Revista Española de Patología* 42.3 (2009): 167-181.
- [3] Ramanujam, Nirmala. "Fluorescence spectroscopy in vivo." *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation* (2006).
- [4] S. Andersson-Engels, J. Johansson, U. Stenram, K. Svanberg, S. Svanberg, "Malignant Tumor and Atherosclerotic Plaque Diagnosis Using Laser Induced Fluorescence", *IEEE J. Quant. Electron.*, 26(12), 2207–2217 (1990).
- [5] S. Andersson-Engels, J. Johansson, K. Svanberg, S. Svanberg, "Fluorescence Imaging and Point Measurements of Tissue: Applications to the Demarcation of Malignant Tumors and Atherosclerotic Lesions from Normal Tissue", *Photochem. Photobiol.*, 53(6), 807–814 (1991).
- [6] S. Andersson-Engels, B.C. Wilson, "In Vivo Fluorescence in Clinical Oncology: Fundamentals and Practical Issues", *J. Cell. Pharmacol.*, 3, 66–79 (1992).
- [7] T.G. Papazoglou, "Malignancies and Atherosclerotic Plaque Diagnosis – Is Laser Induced Fluorescence Spectroscopy the Ultimate Solution?", *J. Photochem. Photobiol. B: Biol.*, 28, 3–11 (1995).
- [8] I.J. Bigio, J.R. Mourant, "Ultraviolet and Visible Spectroscopies for Tissue Diagnostics: Fluorescence Spectroscopy and Elastic-scattering spectroscopy", *Phys. Med. Biol.*, 42, 803–814 (1996).
- [9] S. Andersson-Engels, C.A. Klinteberg, K. Svanberg, S. Svanberg, "In Vivo Fluorescence Imaging for Tissue Diagnostics", *Phys. Med. Biol.*, 42, 815–824 (1997).
- [10] R. Richards-Kortum, E. Sevick-Muraca, "Quantitative Optical Spectroscopy for Tissue Diagnosis", *Annu. Rev. Phys. Chem.*, 47, 555–606 (1996).
- [11] G.A. Wagnieres, W.M. Star, B.C. Wilson, "In Vivo Fluorescence Spectroscopy and Imaging for Oncological Applications", *Photochem. Photobiol.*, 68(5), 603–632 (1998).
- [12] E.B. Hanlon, I. Itzkan, R.R. Dasari, M.S. Feld, R.J. Ferrante, A.C. McKee, D. Lathi, N.W. Kowall, "Nearinfrared Fluorescence Spectroscopy Detects Alzheimer's Disease InVitro", *Photochem. Photobiol.*, 70(2), 236–242 (1999).
- [13] L.T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J.M. Crawford, M.S. Feld, "Observation of Periodic Fine Structure in Reflectance from Biological Tissue: a New Technique for Measuring Nuclear Size Distribution", *Phys. Rev. Lett.*, 80(3), 627–630 (1998).
- [14] S.P. Gopinath, C.S. Robertson, R.G. Grossman, B. Chance, "Near-infrared Spectroscopic Localization of Intracranial Hematomas", *J. Neurosurg.*, 79, 43–47 (1993).

- [15]. W.J. Levy, S. Levin, B. Chance, "Near-infrared Measurement of Cerebral Oxygenation: Correlation with Electroencephalographic Ischemia During Ventricular Fibrillation", *Anesthesiology*, 83(4), 738–746 (1995).
- [16]. S. Nioka, Q. Luo, B. Chance, "Human Brain Functional Imaging with Reflectance CWS", *Adv. Exp. Med. Biol.*, 428, 237–242 (1997).
- [17]. A. Villringer, B. Chance, "Non-invasive Optical Spectroscopy and Imaging of Human Brain Function", *Trends Neurosci.*, 20(10), 435–442 (1997).
- [18]. D.A. Boas, A.G. Yodh, "Spatially Varying Dynamical Properties of Turbid Media Probed with Diffusing Temporal Light Correlation", *J. Opt. Soc. Am. A*, 14, 192–215 (1997)
- [19] Turro, Nicholas J., Vaidhyanathan Ramamurthy, and Juan C. Scaiano. "Book Review Modern Molecular Photochemistry of Organic Molecules." (2012): 1033
- [20] Stedmon, Colin A., Stiig Markager, and Rasmus Bro. "Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy." *Marine Chemistry* 82.3-4 (2003): 239-254.
- [21] Itoh, Takao. "Fluorescence and phosphorescence from higher excited states of organic molecules." *Chemical reviews* 112.8 (2012): 4541-4568.
- [22] Sauer, Markus, Johan Hofkens, and Jörg Enderlein. "Basic principles of fluorescence spectroscopy." *Handbook of fluorescence spectroscopy and imaging: from single molecules to ensembles* (2011): 1-30.
- [23] Lakowicz, Joseph R., ed. *Principles of fluorescence spectroscopy*. Springer Science & Business Media, 2013.
- [24] Rosenthal, Eben, and Kurt R. Zinn. *Optical imaging of cancer*. Springer, 2009.
- [25] [Vo-Dinh, Tuan, ed. *Biomedical photonics handbook: biomedical diagnostics*. CRC press, 2014.
- [26] B. Chance, B. Schoener, R. Oshino, F. Itshak, Y. Nakase, 'Oxidation–Reduction Ratio Studies of Mitochondria in Freeze-trapped Samples', *J. Biol. Chem.*, 254(11), 4764–4771 (1979).
- [27] Kostron, Herwig, and Tayyaba Hasan, eds. *Photodynamic medicine: From bench to clinic*. Vol. 15. Royal Society of Chemistry, 2016.
- [28] F.N. Ghadially, W.J.P. Neish, H.C. Dawkins, 'Mechanisms Involved in the Production of Red Fluorescence of Human and Experimental Tumors', *J. Pathol. Bacteriol.*, 85, 77–92 (1963).
- [29] Brown, Stanley B. "An introduction to spectroscopy for biochemists." *BOOK REVIEWS* 25 (1980).
- [30] F. Koenig, F.J. McGovern, H. Enquist, R. Larne, T.F. Deutsch, K.T. Schomacker, 'Autofluorescence Guided Biopsy for the Early Diagnosis of Bladder Carcinoma', *J. Urol.*, 159(6), 1871–1875 (1998).
- [31] Agostinis, Patrizia, et al. "Photodynamic therapy of cancer: an update." *CA: a cancer journal for clinicians* 61.4 (2011): 250-281.
- [32] Richards-Kortum, Rebecca. "Fluorescence spectroscopy of turbid media." *Optical-Thermal Response of Laser-Irradiated Tissue*. Springer, Boston, MA, 1995. 667-707.

- [33] Keijzer, Marleen, et al. "Fluorescence spectroscopy of turbid media: autofluorescence of the human aorta." *Applied Optics* 28.20 (1989): 4286-4292.
- [34] Richards-Kortum, R., et al. "Spectral diagnosis of atherosclerosis using an optical fiber laser catheter." *American heart journal* 118.2 (1989): 381-391.
- [35] Bigio, Irving J., and Stephen G. Bown. "Spectroscopic sensing of cancer and cancer therapy: Current status of translational research." *Cancer biology & therapy* 3.3 (2004): 259-267.
- [36] Segura Giraldo, Belarmino. "Espectroscopía óptica de fluorescencia aplicada al soporte de diagnóstico médico de precánceres de tejidos de cuello uterino." *Departamento de Ingeniería Eléctrica, Electrónica y Computación* (2009).
- [37] Pu, Yang, et al. "Changes of collagen and nicotinamide adenine dinucleotide in human cancerous and normal prostate tissues studied using native fluorescence spectroscopy with selective excitation wavelength." *Journal of biomedical optics* 15.4 (2010): 047008.
- [38] Andersson-Engels, Stefan, et al. "Fluorescence imaging and point measurements of tissue: Applications to the demarcation of malignant tumors and atherosclerotic lesions from normal tissue." *Photochemistry and photobiology* 53.6 (1991): 807-814.
- [39] Alfano, R., et al. "Human teeth with and without caries studied by laser scattering, fluorescence, and absorption spectroscopy." *IEEE Journal of Quantum Electronics* 20.12 (1984): 1512-1516.
- [40] Ramanujam, Nirmala. "Fluorescence spectroscopy of neoplastic and non-neoplastic tissues." *Neoplasia* 2.1-2 (2000): 89-117.
- [41] Adamko, Darryl, et al. "Variation of metabolites in normal human urine." *Metabolomics* 3.4 (2007): 439-451
- [42] Rajasekaran, Ramu, et al. "Characterization and diagnosis of cancer by native fluorescence spectroscopy of human urine." *Photochemistry and photobiology* 89.2 (2013): 483-491.
- [43] Masilamani, Vadivel, et al. "Cancer detection by native fluorescence of urine." *Journal of biomedical optics* 15.5 (2010): 057003.
- [44] AlSalhi, M., et al. "Detection of cancer by optical analysis of body fluids—a single blind study." *Technology in cancer research & treatment* 10.2 (2011): 145-152.
- [45] Dubayová, Katarína, Jaroslav Kušnír, and L'udmila Podracká. "Diagnostic monitoring of urine by means of synchronous fluorescence spectrum." *Journal of biochemical and biophysical methods* 55.2 (2003): 111-119.
- [46] Masilamani, Vadivel, et al. "Fluorescence spectra of blood and urine for cervical cancer detection." *Journal of biomedical optics* 17.9 (2012): 098001.
- [47] Svanberg, Sune. "Environmental and medical applications of photonic interactions." *Physica Scripta* 2004.T110 (2004): 39.
- [48] Zhu, Timothy C., and Jarod C. Finlay. "The role of photodynamic therapy (PDT) physics." *Medical physics* 35.7Part1 (2008): 3127-3136.
- [49] <https://pubchem.ncbi.nlm.nih.gov/>

- [50] Vicentini, C., et al. "Terapia fotodinámica." EMC-Dermatología 51.3 (2017): 1-8.
- [51] Etcheverry, María E., Miguel A. Pasquale, and Mario Garavaglia. "Photodynamic therapy of HeLa cell cultures by using LED or laser sources." Journal of Photochemistry and Photobiology B: Biology 160 (2016): 271-277.
- [52] Klinteberg, Claes, et al. "Kinetic fluorescence studies of 5-aminolaevulinic acid-induced protoporphyrin IX accumulation in basal cell carcinomas." Journal of Photochemistry and Photobiology B: Biology 49.2-3 (1999): 120-128.
- [53] R. F. Revenga Arranza, «Las verrugas,» Medicina Integral. Elsevier, vol. 37, nº 9, pp. 375-405, 2001.
- [54] <https://www.mayoclinic.org/es-es/diseases-conditions/oral-lichen-planus/symptoms-causes/syc-20350869>
- [55] Asociación Española de Patología Cervical y Colposcopia, divulgación. CONDILOMAS ACUMINADOS, disponible en <http://www.aepcc.org/condilomas/>
- [56] Corti, María Agustina, Mario J. Garavaglia, and Horacio Poteca. "Case report: optical biopsy in HPV6 lesion." Imaging in Medicine 10 (2018).
- [57] Fazzini, Romeo. "Tumores de Glándulas Salivales." Facultad de Medicina de Buenos Aires 1.1 (2000).
- [58] Escrig-Sos, Javier, David Martínez-Ramos, and Juan Manuel Miralles-Tena. "Pruebas diagnósticas: nociones básicas para su correcta interpretación y uso." Cirugía Española 79.5 (2006): 267-273.
- [59] De Burgueño, M. J., J. L. García-Bastos, and J. M. González-Buitrago. "Las curvas ROC en la evaluación de las pruebas diagnósticas." Med Clin (Barc) 104.17 (1995): 661-70.
- [60] Zweig, Mark H., and Gregory Campbell. "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine." Clinical chemistry 39.4 (1993): 561-577.
- [61] Bro, Rasmus, and Age K. Smilde. "Principal component analysis." Analytical Methods 6.9 (2014): 2812-2831.
- [62] Jolliffe, Ian T., and Jorge Cadima. "Principal component analysis: a review and recent developments." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374.2065 (2016): 20150202.
- [63] Tauler, Roma. "Multivariate curve resolution applied to second order data." Chemometrics and intelligent laboratory systems 30.1 (1995): 133-146.
- [64] Balakrishnama, Suresh, and Aravind Ganapathiraju. "Linear discriminant analysis-a brief tutorial." Institute for Signal and information Processing 18 (1998): 1-8.
- [65] Bro, Rasmus. "PARAFAC. Tutorial and applications." Chemometrics and intelligent laboratory systems 38.2 (1997): 149-172.
- [66] Andersen, Charlotte Møller, and R. Bro. "Practical aspects of PARAFAC modeling of fluorescence excitation-emission data." Journal of Chemometrics: A Journal of the Chemometrics Society 17.4 (2003): 200-215.

- [67] Li, Shousong, and Paul J. Gemperline. "Eliminating complex eigenvectors and eigenvalues in multiway analyses using the direct trilinear decomposition method." *Journal of chemometrics* 7.2 (1993): 77-88.
- [68] Sands, Richard, and Forrest W. Young. "Component models for three-way data: An alternating least squares algorithm with optimal scaling features." *Psychometrika* 45.1 (1980): 39-67.
- [69] Murphy, Kathleen R., et al. "Fluorescence spectroscopy and multi-way techniques. PARAFAC." *Analytical Methods* 5.23 (2013): 6557-6566.
- [70] Harshman, Richard A., and Margaret E. Lundy. "PARAFAC: Parallel factor analysis." *Computational Statistics & Data Analysis* 18.1 (1994): 39-72.
- [71] Harshman, Richard A., and Margaret E. Lundy. "The PARAFAC model for three-way factor analysis and multidimensional scaling." *Research methods for multimode data analysis* 46 (1984): 122-215.
- [72] Atif, M., et al. "A study for the detection of kidney cancer using fluorescence emission spectra and synchronous fluorescence excitation spectra of blood and urine." *Photodiagnosis and photodynamic therapy* 23 (2018): 40-44.
- [73] Masilamani, Vadivel, et al. "Fluorescence spectra of blood and urine for cervical cancer detection." *Journal of biomedical optics* 17.9 (2012): 098001.
- [74] AlSalhi, M., et al. "Detection of cancer by optical analysis of body fluids—a single blind study." *Technology in cancer research & treatment* 10.2 (2011): 145-152.
- [75] Lualdi, Manuela, et al. "Natural fluorescence spectroscopy of human blood plasma in the diagnosis of colorectal cancer: feasibility study and preliminary results." *Tumori Journal* 93.6 (2007): 567-571.
- [76] Anwer, Ayad G., et al. "Distinctive autofluorescence of urine samples from individuals with bacteriuria compared with normals." *Clinica chimica acta* 401.1-2 (2009): 73-75.
- [77] (Saldaña, Manuel Romero. "Pruebas de bondad de ajuste a una distribución normal." *Revista Enfermería del trabajo* 6.3 (2016): 114.
- [78] Gamboa, Nadia. "El empleo de la derivada en la espectroscopía ultravioleta-visible." *Revista de Química* 5.2 (1991): 179-195.
- [79] Press, William H., and Saul A. Teukolsky. "Savitzky-Golay smoothing filters." *Computers in Physics* 4.6 (1990): 669-672.
- [80] Kitamura, Keisuke, and Keiichiro Hozumi. "Effect of the degree of polynomials in the Savitzky—Golay method for calculation of second-derivative spectra." *Analytica chimica acta* 172 (1985): 111-118.
- [81] Leiner, Marc JP, Michael R. Hubmann, and Otto S. Wolfbeis. "The total fluorescence of human urine." *Analytica Chimica Acta* 198 (1987): 13-23.
- [82] Kušnír, Jaroslav, et al. "Concentration matrices—solutions for fluorescence definition of urine." *Analytical letters* 38.10 (2005): 1559-1567.

[83] Sandeepa, Perinchery M., et al. "Factors influencing the UV autofluorescence of human urine." International Conference on Optics and Photonics. 2009.